Journal of
Experimental Psychology

VOL. 29, NO. 2 AUGUST, 1941

THREE ATTEMPTS TO SECURE PUPILLARY CONDITIONING TO AUDITORY STIMULI NEAR THE ABSOLUTE THRESHOLD

BY ERNEST R. HILGARD, JOSEPH MILLER,[1] AND JOHN A. OHLSON

Stanford University

Results from experimental attempts to condition the pupillary response in man are in an unfortunate state of confusion. The early success of Cason (3) was repeated and extended by Hudgins (4), and Kotliarevsky (5). In several experiments Steckle and Renshaw (9), Steckle (8), and Zabarenko (11) have failed to repeat the results obtained by Hudgins. A report by Baker (2) added considerations which, if confirmed, would be of utmost importance to conditioning theory. He found that pupillary responses could be conditioned to subthreshold auditory stimuli consistently in three trials, the responses thus conditioned not being subject to extinction by non-reinforcement, and persisting for as long as two years. Responses conditioned to supra-threshold sounds were established with much greater difficulty, and were less stable. Data supporting the original conclusions have been added by Metzner and Baker (6). A failure to confirm Baker's results has been reported by Wedell, Taylor, and Skolnick (10). The three experiments to be described were designed to repeat Baker's work.[2] Because of the specific arrangements involved, they cannot be considered as tests of the experiments of Cason or of Hudgins.

[1] This writer participated in these experiments during the course of tenure of a National Research Council Fellowship and wishes to acknowledge his indebtedness to the National Research Council Fellowships Board.

[2] Dr. Baker has been very helpful in facilitating the repetition of his investigation through furnishing photographs and drawings of his apparatus and through giving supplementary information about his procedure. Several suggestions obtained through correspondence with him have been incorporated in this report.

Apparatus and Procedure

Auditory Apparatus and Threshold Determination.—The apparatus used by Baker and his method of threshold determination were described in an earlier paper, Baker (1). In our first and second experiments we substituted an oscillator producing relatively pure tones for the noise-producing audiometer used by him, but in our third experiment we attempted to duplicate his apparatus as faithfully as possible. The beat-frequency oscillator used was built specially for the Stanford Laboratory, and provides a continuous range of intensities at each of several frequencies. The frequency adopted for experimental use was 1900 cycles. For the production of noises, a motor running at 1725 R.P.M. was fitted with a commutator which interrupted the direct current through the headphones once each revolution. The noise was attenuated by a Seashore audiometer, the steps of which, in common with the scale which appears on it and in common with Baker's practice, are described as "j.n.d.'s." The oscillator is constructed with a switch in the circuit inserted in such a manner that there is no click at the onset or termination of the tone. The whole arrangement was much freer of extraneous sounds than that in which the noises were presented. In the case of the noise stimuli, the running motor was in the same room with the subject and a spring key on the audiometer was used to turn the sound on and off, duplicating as nearly as possible Baker's conditions.

The stimuli presented in the determination of the auditory threshold were patterns of tones or noises, either a dot-dash or a dash-dot. The timing, by means of a stopwatch, was under one second for the dot, one second for the space, and two seconds for the dash. The subject was asked to report each time either dot-dash or dash-dot, guessing if he heard nothing. After each report he assigned a value for the degree of confidence of his judgment, '2' for great confidence ('perfectly sure'), '1' for slight confidence, '0' for no confidence whatever ('pure guess'). The instructions were copied from those reported by Baker. While the patterns of dot-dash and dash-dot followed one another in random order, we used at different times two different procedures with respect to the order of intensities. The first procedure, that of Baker, is a modification of the method of limits, in which intensity changes at each trial. In his procedure, ascending and descending orders alternated, the length of series up or down varying, but never being over ten audiometer steps. The second procedure, the method of serial groups, consists in securing a group of judgments (ten in our experiments) before changing from one intensity to another. The two procedures did not produce differences important enough to affect the interpretations of our disagreement with Baker's findings.

We were unable to secure a correspondence between the percentage of correct judgments and the degree of confidence in them which compared at all with Baker's data, although our own results with tones and noises, under quite varied environmental circumstances, were relatively self-consistent. Our results, along with those of Baker, are presented in Table 1. Examination of the columns labelled 'Percent Correct' shows a fair correspondence among the three sets of values. It appears that the steps used by Baker may be slightly smaller than those used by us, so that a range of seven steps in our values corresponds to a range of eight in his. But the columns labelled 'Percent Guesses' show striking disagreement between our results and Baker's. While his subjects report 100 percent guessing while making 86 percent correct responses, our subjects do not report 100 percent guessing until their responses are practically at a chance level. The sudden jump for Baker's subjects from 34 percent guessing at 87 percent correct to 100 percent guessing at 86 percent correct is surprising, in view of the gradual increases in guessing shown by our subjects between 30 and 100 percent. The pooling of results from different subjects might account in part for these differences, but we have been unable to select individual subjects or pool our data in any other way which would make the results approach those reported by Baker. The source of disagreement eludes us, because we have tried alternative instructions encouraging more guesses, alternative orders of stimuli, quiet and noisy conditions of experimentation. Our results, while not altogether self-consistent in terms of classical psychophysics, never have approximated those which we were seeking to duplicate.

That guesses did not reach 100 percent until responses were at chance need not be at variance with the known fact that under proper circumstances a subject can respond to subthreshold stimuli, as reported, for example, by J. G. Miller (7). From the psychophysical function we may predict that the 100 percent point of guessing will tend to fall out beyond the point of substantial amounts of guessing. Our subjects are actually reporting at above chance values correct when their responses are predominantly guesses, although not 100 percent guesses.

Because of this failure to support one of the initial requirements of the experiment we were attempting to repeat, an arbitrary decision was necessary. The value called subthreshold and chiefly used in conditioning by Baker was that at a value of −2 audiometer steps in Table 1, which for his subjects yielded 100 percent guessing at 69 percent correct. In our experiments 1 and 2 we adopted as a conditioning value that audiometer position which yielded for the subject most nearly 70 percent correct responses, regardless of the proportion of guesses. It appeared to us that this percentage correct would duplicate as best we could the intensity used by Baker. In Experiment 3 we adopted a suggestion from the experiment of Wedell, Taylor, and Skolnick (10).[3] They attempted to condition at each of several intensities, beginning distinctly below the threshold, in order to see at which intensity conditioning would appear. Because of the uncertainty of the intensity level at which Baker had worked, in our third experiment we used values yielding approximately 50, 70, and 90 percent correct judgments, lest in using the 70 percent alone we had been working at an intensity of sound too different from that used by Baker.

TABLE 1

PERCENTAGE OF CORRECT JUDGMENTS AND PERCENTAGE OF PURE GUESSES WITH NOISES AND TONAL STIMULI

Audiometer Steps (Arbitrary Scale*)	Noise 4 Subjects (Baker (1))		Tone 10 Subjects (Exp. 2)		Noise 21 Subjects (Exp. 3)	
	Percent Correct	Percent Guesses	Percent Correct	Percent Guesses	Percent Correct	Percent Guesses
+3	94	14	99	7	100	0
+2	91	22	91	7	98	2
+1	87	34	84	18	96	6
0	86	100	87	33	87	16
−1	82	100	81	42	78	31
−2	69	100	68	66	64	53
−3	62	100	61	74	56	80
−4	54	100	47	100	52	100

* The zero value is placed, in accordance with Baker's practice, at that value at and below which judgments were given as pure guesses in his study. The audiometer settings at which corresponding percentages of correct judgments were secured in Experiments 2 and 3 were arbitrarily designated as zero. Steps are in 'j.n.d.'s' on Seashore's audiometer in the case of noises, and in decibels in the case of tones.

Visual Stimuli and Pupillary Measurement.—The illumination change which constituted the unconditioned stimulus appeared on a milk glass before the eyes of the subject. The glass was illuminated from behind by a 60 watt, 120 volt lamp, as described by Baker. Low, high, and intermediate illuminations were used, the values being determined according to Baker's practice. The low value, 3.5 foot-candles as measured by a Weston Photometer placed at the position of the subject's eyes, was found to correspond in our experiment to a brightness of 78 ml as measured with a Macbeth Illuminometer. Our intermediate value represented a brightness of 220 ml, and our high value 1125 ml. Since the diffusion properties of the opal flash glass used by Baker are unknown, brightness cannot be inferred precisely from his reported foot-candle values, but we believe that our light conditions approximated his.

The subject fixated binocularly at either of two distances. Near fixation, corresponding to that used by Baker, was upon a spot in the center of the lens of the telescope before the right eye at a distance of 15 cm. Baker's subjects fixated binocularly the image of the right eye in the lens at a distance of 7 inches. For far fixation, a mirror was arranged just above the illuminated window which reflected the image of a fixation cross, so that the cross appeared to the subject

[3] The suggestion was contained in a letter from Dr. Wedell dated May 22, 1939, received after we had already completed our Experiments 1 and 2.

to be straight ahead at 100 cm. Binocular fixation is more easily maintained at this greater distance.

Pupillary size was recorded on a polygraph by means of a manually operated system attached to a telescope through which the pupil was viewed. There were a pair of vertical cross-hairs appearing in the field of vision of the recorder, these being kept at the edges of the pupil by turning a knurled knob. The turning of the knob activated the recording system. The device was so calibrated that a magnification of exactly 10 : 1 was secured on the recording paper. The magnitude of pupillary changes dealt with may be seen in Table 2.

TABLE 2

EFFECT ON PUPILLARY SIZE OF NEAR AND FAR BINOCULAR FIXATION

Near fixation upon a white spot in center of lens at distance of 15 cm; far fixation upon small illuminated cross at 100 cm. In Experiments 1 and 2 there were rests between trials, with new fixation each trial; in Experiment 3 fixation was prolonged for several trials.

	Pupillary Diameter in mm					
	New Fixations Each Trial				Prolonged Fixation	
	Constriction		Dilation		Constriction	
	Far (11 S's)	Near (10 S's)	Far (10 S's)	Near (11 S's)	Far (10 S's)	Near (9 S's)
Prior to illumination change	5.5	3.9	3.9	3.6	5.0	5.1
During changed illumination	3.7	3.1	4.9	4.3	3.2	3.3
Change in diameter of pupil	−1.8	−0.8	+1.0	+0.7	−1.8	−1.8

When new fixations were made each trial, pupillary diameter was larger initially with far fixation and the change was slightly greater as the light increased or decreased. With prolonged fixation, however, there appeared to be an adaptation to the different distances of binocular fixation, and mean pupillary diameters were essentially alike for far and near fixation. It is quite possible that changes in fixation, related to expected changes in stimuli within conditioning may be important sources of error in conditioned pupillary experiments.

The corresponding values for pupillary changes from Baker indicate an error in his calibration, so that his reported changes are approximately 5 times too large.[4] This is a systematic error, however, which in no way affects the interpretations of his data.

The Conditioning Procedure.—While some modifications in the order of trials were introduced in the different experiments, the basic pattern for a single trial remained constant throughout. Each trial endured for 36 seconds, divided into four segments of 9 seconds each. The first and last segments were always control periods, the pupillary diameter and its changes being recorded during these periods of no stimulus changes. There were three types of trial: conditioning, test, and control. In a conditioning trial, the initial 9 seconds of control were followed by 9 seconds of sound alone (the conditioned stimulus). Then the illumination change (unconditioned stimulus) occurred and endured for 9 seconds, the sound continuing. Light and sound were terminated together at the end of this period, and the trial continued for 9 more seconds with the light at its original prestimulus intensity. In a test trial, following the 9 seconds of control the sound began and continued for 18 seconds without illumination change, then ceased for the

[4] Wedell, Taylor, and Skolnick (10) report that Baker's values should be divided by 2, instead of by 5. Our factor was arrived at on the assumption that maximum constriction to illumination increase and maximum dilation to illumination decrease should have been alike in Baker's experiments and ours. If our illumination values depart from his, the factor may be slightly in error.

remaining 9 seconds of the trial. In the control trials no stimuli were presented within the 36 seconds. The basic temporal arrangement is that of Baker, 9 seconds of sound followed by 9 seconds of light and sound, so that the procedure provides a fair test of his practice. Stimuli were presented manually by one experimenter, while another traced the pupillary diameter. The one tracing had no further duties at the time. The experimenters rotated positions with different subjects. Each of the three authors served in each experimental position.[5]

Time marks appeared on the record each 3 seconds, and the pupillary diameter was measured at each of the 3-second positions throughout the trial. In the later tables, mean values of changes in pupillary diameter are reported for each of these times from the beginning of the trial. By using this standard of measurement, it was hoped that any distortion in reading the record due to judging the presence or absence of a response might be ruled out. This is essentially Baker's practice. While Baker reported values for each second, values for each 3 seconds would suffice to confirm his findings, if responses such as his were to appear on the records.

The number of trials and the conditions of rest and fixation will be described separately under each of the three experiments to be reported.

Experiment 1. Attempted Conditioning with Tonal Stimuli, Using Far Fixation

Auditory thresholds to a tone of 1900 cycles were determined on the first day, as indicated above. Conditioning was then undertaken on the second day of experimentation. After several preliminary test trials, there were 8 conditioning trials in two sets of 4 each with two control trials interspersed. These were followed in turn by 5 test trials (constituting a brief extinction series). Between each bank of 5 trials the subject removed his head from the rest. Then he was instructed to take position, and shortly before each trial began was given instructions to fixate. At the end of each trial, the subject was permitted to relax, but kept his head in the rest. The pupil was recorded only within a period beginning shortly before and ending shortly after each trial. This differs from Baker's practice, but the procedure was adopted in the early experiments in order to reduce fatigue and to minimize the spontaneous fluctuations of the pupil. During conditioning, the light changed from low to high brightness for attempted conditioned constriction and from high to low brightness for attempted conditioned dilation. The test trials at the beginning and end were conducted with the light at mid-position, again in accordance with Baker's procedure. Because of these changes in the initial light value, and hence in initial pupillary diameter, changes are expressed in the tables which follow as departures from the value at the end of 9 seconds, that is, at the end of the control period and at the onset of the conditioned stimulus. The mean absolute diameters are given in the footnotes to each table.

The results for attempted conditioned dilation are given in Table 3. Conditioning, if present, should be evidenced during the period that the tone alone is sounding, particularly at the end of this period

[5] In Experiments 1 and 2, Hilgard and Ohlson were the experimenters; in Experiment 3, Hilgard and Miller.

(the 18th second). The values for conditioning are to be compared with the control trials, while the values within the final test should be compared with the preliminary test trials. The departure at the end of the 18th second within conditioning is in the right direction, but the mean amount is only a fraction of that reported by Baker, and is counterbalanced by departure in the wrong direction in the

TABLE 3

PUPILLARY CHANGES WITHIN ATTEMPTED CONDITIONED DILATION, EXPERIMENT 1

Conditioned stimulus, tone of 1900 cycles; far binocular fixation; rests between trials; mean values from 10 subjects

	Pupillary Change in Millimeters *				
	Seconds from Beginning of Trial	Conditioning		Test	
		Mean, 2 Control Trials †	Mean, Second 4 Conditioning Trials	Mean, 2 Preliminary Trials ‡	Mean, 5 Final Test Trials
Control period (no stimuli)	3	+.10	+.07	+.13	+.13
	6	+.13	+.03	−.01	+.03
	9	.00	.00	.00	.00
Tone only	12	+.01	−.04	−.03	+.02
	15	+.02	−.08	−.02	−.03
	18	−.04	+.08	−.02	−.07
Tone only in test trials; tone + light in conditioning	21	+.01	+.64	−.06	−.06
	24	+.08	+.88	−.13	−.07
	27	+.08	+1.02	−.12	−.03
Control period (no stimuli)	30	+.03	−.50	−.09	−.11
	33	+.10	−.28*	−.16	−.02
	36	+.01	−.22	−.14	−.08

* The pupillary diameter at 9 secs. was in each case taken as zero, and changes calculated from that value; absolute value 3.9 mm prior to attempted conditioned dilation, 4.5 mm within test series.

† Control trials interspersed among conditioning trials. No tone.

‡ Trials prior to attempted conditioning. Tone only. Light at testing intensity.

final test trials. Statistical tests (to be presented later, in Table 8) show the amount of conditioning to be lacking in significance, so that results for dilation must be considered negative.

Table 4 contains the corresponding data for attempted conditioned constriction. All circumstances of experimentation were like those of attempted conditioned dilation except that during conditioning trials the light began at low values and was increased to produce the unconditioned pupillary change. The values for the 18th second show nearly as much constriction in the control trials as in conditioning and more in preliminary than in final test trials. Like the results for dilation, these fail to demonstrate conditioning.

There were three chief differences in conditions in Experiment 1 from those of Baker's experiment, any one of which might have been responsible for the differences in findings. These were (1) the use of tones instead of noises as conditioned stimuli; (2) the use of rests between trials instead of continuous fixation; (3) the use of a far rather than a near fixation point. Because of the wide fluctuations

TABLE 4

PUPILLARY CHANGES WITHIN ATTEMPTED CONDITIONED CONSTRICTION, EXPERIMENT 1
Conditioned stimulus, tone of 1900 cycles; far binocular fixation; rests between trials; mean values from 11 subjects

	Pupillary Change in Millimeters *				
	Seconds from Beginning of Trial	Conditioning		Test	
		Mean, 2 Control Trials †	Mean, Second 4 Conditioning Trials	Mean, 2 Preliminary Trials ‡	Mean, 5 Final Test Trials
Control period (no stimuli)	3	+.13	+.01	+.09	+.12
	6	+.07	+.02	+.03	+.11
	9	.00	.00	.00	.00
Tone only	12	+.01	.00	−.01	+.02
	15	−.06	−.07	−.05	+.04
	18	−.06	−.07	−.12	−.02
Tone only in test trials; tone+light in conditioning	21	−.12	−2.04	−.21	−.05
	24	−.12	−1.80	−.12	−.04
	27	−.04	−1.76	−.07	−.11
Control period (no stimuli	30	−.04	−.84	−.11	−.11
	33	−.16	−.50	−.16	−.10
	36	−.18	−.36	−.17	−.13

* The pupillary diameter at 9 secs. was in each case taken as zero, and changes calculated from that value; absolute value 5.5 mm prior to attempted conditioned constriction, 4.7 mm within test series.

† Control trials interspersed among conditioning trials. No tone.

‡ Trials prior to attempted conditioning. Tone only. Light at testing intensity.

in pupillary diameter with changes in binocular fixation, it was decided to keep other conditions constant but to change that one for the next group of subjects.

EXPERIMENT 2. ATTEMPTED CONDITIONING WITH TONAL STIMULI, USING NEAR FIXATION

The order of trials, the instructions and the stimuli used, were alike in Experiments 1 and 2. The only change was the substitution of the 15 cm fixation for that of 100 cm. Binocular fixation at such a short distance involves considerable strain, and it was felt that perhaps loss of fixation and its recovery when light changes occurred

(or when they were expected) might be responsible for the positive results which Baker secured and the negative results which we secured with far fixation.

The results for attempted conditioned dilation are given in Table 5. Making the usual comparison of pupillary changes at the 18th second, when conditioning should be most in evidence, it may be

TABLE 5

PUPILLARY CHANGES WITHIN ATTEMPTED CONDITIONED DILATION, EXPERIMENT 2

Conditioned stimulus, tone of 1900 cycles; near binocular fixation; rests between trials; mean values from 11 subjects

	Pupillary Change in Millimeters *				
	Seconds from Beginning of Trial	Conditioning		Test	
		Mean, 2 Control Trials †	Mean, Second 4 Conditioning Trials	Mean, 2 Preliminary Trials ‡	Mean, 5 Final Test Trials
Control period (no stimuli)	3	.00	+.01	+.12	+.05
	6	+.03	−.02	−.01	+.01
	9	.00	.00	.00	.00
Tone only	12	+.01	+.04	+.08	−.04
	15	+.08	+.04	+.07	−.04
	18	+.12	+.06	+.11	−.07
Tone only in test trials; tone+light in conditioning	21	+.02	+.61	+.11	−.11
	24	−.03	+.78	+.07	−.05
	27	+.05	+.79	+.13	−.08
Control period (no stimuli)	30	+.06	−.40	+.15	−.09
	33	+.11	−.20	+.15	−.06
	36	+.03	−.10	+.14	−.07

* The pupillary diameter at 9 secs. was in each case taken as zero, and changes calculated from that value; absolute value 3.5 mm prior to attempted conditioned dilation, 3.9 during test series.

† Control trials interspersed among conditioning trials. No tone.

‡ Trials prior to attempted conditioning. Tone only. Light at testing intensity.

seen that there is more dilation in the control trials and preliminary tests than there is in the conditioning or final test trials. The results are again disappointingly negative. It should be pointed out that here, as in the earlier tables, the stability of the baseline of pupillary diameter and the change introduced by the unconditioned stimulus are sufficient that failure of conditioning cannot be attributed to lack of sensitivity of the recording devices or to the concealing effect of spontaneous pupillary changes.

Results for attempted conditioned constriction as given in Table 6 show that at the 18th second both the conditioning and final test trials depart from the controls and preliminary trials in the right

direction. If the mean values were larger and reliable, they would confirm Baker's findings. However, the corresponding constriction reported by him is 3.3 mm, which, corrected, is 0.66 mm, or several times the value obtained by us. The constriction continues in the final control periods of the test trials (seconds 30 to 36), instead of returning to the baseline, so that the constriction is probably not to

TABLE 6

PUPILLARY CHANGES WITHIN ATTEMPTED CONDITIONED CONSTRICTION, EXPERIMENT 2

Conditioned stimulus, tone of 1900 cycles; near binocular fixation; rests between trials; mean values from 10 subjects

	Pupillary Change in Millimeters *				
	Seconds from Beginning of Trial	Conditioning		Test	
		Mean, 2 Control Trials †	Mean, Second 4 Conditioning Trials	Mean, 2 Preliminary Trials ‡	Mean, 5 Final Test Trials
Control period (no stimuli)	3	+.07	−.01	−.11	+.05
	6	+.07	+.02	−.01	+.02
	9	.00	.00	.00	.00
Tone only	12	.00	−.03	+.04	−.06
	15	+.02	−.07	−.07	−.07
	18	+.11	−.10	−.10	−.11
Tone only in test trials; tone+light in conditioning	21	+.15	−1.13	−.13	−.10
	24	+.05	−.96	−.12	−.14
	27	+.02	−.82	−.14	−.11
Control period (no stimuli)	30	+.01	−.15	−.08	−.09
	33	+.03	−.07	−.09	−.10
	36	+.04	−.03	−.15	−.09

* The pupillary diameter at 9 secs. was in each case taken as zero, and changes calculated from that value; absolute value 3.9 mm prior to attempted conditioned constriction, 3.6 mm during test series.

† Control trials interspersed among conditioning trials. No tone.

‡ Trials prior to attempted conditioning. Tone only. Light at testing intensity.

be attributed to the presence of the auditory stimulus. In agreement with the findings in other portions of the experiments, the changes which might be indicative of conditioning are not significant statistically.

The data presented in Tables 5 and 6 suffice to show that the change from far fixation to near fixation was not enough to produce the results found by Baker. It was therefore necessary to approximate his conditions even more closely in order to determine, if possible, the circumstances requisite for reproducing his findings.

Experiment 3. Attempted Conditioning with Three Intensities of Noise Stimuli

Two further changes were introduced in Experiment 3 in order to make the conditions conform more closely to Baker's. Noise was substituted for tone by means of the arrangement previously described. Since our threshold calculations did not agree with his, three different noise levels were used as conditioned stimuli, one at which the percentage of correct judgments was near chance, and confidence expressed as nearly 100 percent guess, a second at judgments averaging 70 percent correct, and a third at judgments averaging 90 percent correct, as determined for each individual subject by a procedure as much like Baker's as we were able to make it. The use of noise at three levels was the first major change. The second was the elimination of rests between successive trials, fixation being maintained for 6 consecutive trials before a rest was introduced. The pupillary diameter was recorded continuously, although measurements are reported as before only for the 36 seconds officially considered one trial. Near and far fixation were alternated, each subject serving with one or the other, but no significant differences were found, and the results are presented for all as one group. Since constriction is somewhat more satisfactory to work with because of the sharp onset of the unconditioned response, all of the subjects were run under this condition.

A minor change in procedure was also added, in that the experimenter doing the tracing signalled to the operating experimenter when the pupil appeared to be relatively stable, hence when a trial should begin. Since Baker had acted as his own tracer and operator, it was thought that perhaps the more active participation of the tracer might produce some effect. Each bank of trials consisted of four reinforcements followed by one test and one control trial. In order to prevent the tracer from being too much influenced by his own suggestion, test and control trials were alternated without the tracer's knowledge. While he therefore signalled that a trial should start, he did not know what kind of a trial it would be, except during the reinforcement trials, and he did not know precisely when those would begin. One such series was run at the lowest chosen sound intensity, then a rest was introduced. A second series followed with the sound of middle intensity, then there was another rest. A third similar series with the loudest sound completed the conditioning. In Table 7 the results are presented separately for the three different sound intensities.

Results with the lowest intensity of auditory stimulus, at which judgments were near to the 50 percent threshold, and confidence

TABLE 7

PUPILLARY CHANGES WITHIN ATTEMPTED CONDITIONED CONSTRICTION: EXPERIMENT 3

Conditioned stimulus, noise produced in headphones by interruption of direct current; fixation prolonged through 6 trials without rest or ready signals; 19 subjects, 10 with far, 9 with near binocular fixation; control and test trials from 16 of these subjects, low and medium intensities, 15 with high intensity noise

	Seconds from Beginning of Trial											
	3	6	9	12	15	18	21	24	27	30	33	36
	Control Period (No Stimuli)			Sound Only; No Stimuli in Control Trial			Sound in Test Trial; Sound-Light in Conditioning; No Stimuli in Control			Control Period (No Stimuli)		
Low intensity noise												
Conditioning[2]	+.04	−.01	.00*	−.13	−.16	−.24	−2.05	−1.90	−1.77	−.52	−.26	−.10
Control[2]	+.09	+.07	.00	+.03	−.04	−.11	−.17	−.07	−.04	−.19	−.17	−.24
Test[3]	−.05	−.02	.00	−.13	−.32	−.26	−.26	−.31	−.32	−.29	−.43	−.35
Medium intensity noise												
Conditioning[2]	+.01	−.02	.00	−.07	−.13	−.19	−1.79	−1.57	−1.44	−.31	−.17	.00
Control[2]	+.09	+.15	.00	+.02	−.02	+.03	−.02	−.09	−.04	−.06	−.10	−.03
Test[3]	−.09	.00	.00	−.08	−.11	−.10	−.06	−.18	−.15	−.25	−.24	−.16
High intensity noise												
Conditioning[2]	+.03	−.01	.00	+.05	−.03	−.13	−1.85	−1.64	−1.51	−.36	−.17	−.09
Control[2]	−.03	+.03	.00	+.01	+.02	−.04	−.06	.00	−.07	−16	−.20	−.12
Test[3]	−.10	−.04	.00	−.10	−.06	−.05	−.08	−.04	−.10	−.10	−.08	−.16

[1] Conditioning trials, mean of trials 3 and 4.

[2] Control, 1 trial following conditioning.

[3] Test, 1 trial following conditioning.

* Pupillary diameters are calculated from the 9 sec. value as zero; absolute value 4.8 mm prior to attempted conditioning, 4.7 mm in control and test trials for high intensity; 4.7 mm prior to attempted conditioning, 4.5 mm in control and test trials for medium intensity; 5.0 mm prior to conditioning, 4.7 mm in control and test trials for low intensity.

expressed as pure guess, may first be considered. During the period of sound alone there is a progressive constriction shown both in conditioning and test trials to a somewhat greater amount than in the control trials, but the significance of the differences is in neither case acceptable. Although conditioning is not demonstrated, the results do not contradict Baker's findings, since the auditory stimuli, in terms of correct judgments, were much below those he used successfully.

In Table 7 there are also given the results for the stimuli of medium intensity, those interpreted by us as being most nearly the ones used by Baker. At the 18th second the control trials have shown

slight dilation, while both conditioning and test trials show some constriction. In absolute amount, conditioning is shown by a mean value about one-third that expected from Baker's findings, in the test trials about one-fifth that expected, and the means do not meet statistical requirements for significance. It is of interest, however, that of all conditions tried out, this one, nearest that used by Baker, gave the most consistent indication of results in the direction of those he reported.

Finally, the data obtained with the highest intensity of auditory stimulus are also summarized in Table 7. These stimuli were correctly reported in 90 percent of the cases, mostly with a high degree of confidence. Conditioning is not in evidence, as judged by the small differences between results in the control and test trials. Since the sound was audible, control trials should perhaps have been arranged prior to conditioning, to detect any unconditioned pupillary response to sound. Since no difference was obtained within the sound trials and the silent trials at the end, the further precaution was perhaps unnecessary.

Conclusions and Interpretations

A summary of critical comparisons between conditioning and test trials and the appropriate control or preliminary test trials is given in Table 8. The mean difference over the control at the 18th second should demonstrate conditioned dilation or conditioned constriction by significant departures in the expected direction. Of the fourteen comparisons in the table, ten are in the expected direction, four in the reverse direction. Not one of the differences is statistically significant by usual conventions, the most significant value being that for conditioning with the noise of medium intensity in Experiment 3. The mean constriction as compared with the control yields a *t*-value of 2.05, which (for 18 degrees of freedom) means that the probability is above .05 that the result would occur by chance. Since the experimenter's knowledge of stimulus conditions introduces a suggestion factor in the conditioning trials, a more critical comparison is that between control and test trials, since the experimenter was without knowledge of these. With the medium auditory stimulus, the *t*-value, while higher than for the other intensities, is only 1.47, and might readily occur by chance. These unsatisfactory mean findings are to be compared with those of Baker in which *every* subject conditioned promptly, yielding a final form response on all trials beyond the third.

How can the differences be explained? The following items would seem *not* to be important in determining the differences:

TABLE 8

SUMMARY OF CRUCIAL COMPARISONS BETWEEN CONDITIONING, TEST, AND CONTROL TRIALS

In each case the comparison is made on the basis of the change in pupillary diameter between the 9th and 18th second, when the conditioned stimulus was presented alone within conditioning and test trials.

	Change in Pupillary Diameter from that within Control or Preliminary Test Trials (mm)				
	Cases	Mean	Standard Error	't'	Direction of Change
Experiment 1					
Dilation					
Conditioning	10	+.117	.073	1.60	Expected
Test trials	10	−.054	.063	.86	Reverse
Constriction					
Conditioning	11	−.007	.087	.08	Expected
Test trials	11	+.103	.076	1.35	Reverse
Experiment 2					
Dilation					
Conditioning	11	−.061	.089	.69	Reverse
Test trials	11	−.185	.120	1.54	Reverse
Constriction					
Conditioning	10	−.212	.128	1.66	Expected
Test trials	10	−.013	.094	.14	Expected
Experiment 3					
Constriction					
(*a*) Low auditory intensity					
Conditioning	19	−.137	.162	.85	Expected
Test trials	16	−.150	.177	.85	Expected
(*b*) Medium intensity					
Conditioning	19	−.221	.108	2.05	Expected
Test trials	16	−.131	.089	1.47	Expected
(*c*) High intensity					
Conditioning	17*	−.012	.136	.09	Expected
Test trials	15	−.012	.124	.10	Expected

* There were 19 subjects conditioned, but control records were missing for two of these. The number of cases in Experiment 3 is in each instance the total number for whom the given comparison is possible with the data available.

1. Noises instead of tones. We used a range of noises as well as of tones, as did also Wedell, Taylor, and Skolnick. The range probably corrects for any ambiguity involved in reproducing Baker's threshold measurement.

2. Continuous fixation instead of rests between trials. We used both.

3. Near instead of far fixation. While we used both, there is still the chance that lack of *any* adequate fixation point in Baker's situation may have been important. His subjects tried to fixate the image of their own eyes in the lens. With changes in illumination such as those used in the experiment it is impossible to see this image continuously.[6]

[6] It is to be noted that Hudgins (4) also lacked an adequate fixation point.

4. Subjectivity of recording. The three of us who served as tracers felt that there was genuine difficulty in giving an objective record of the details of pupillary change, but there was little question about noting the presence or absence of changes of the magnitude reported by Baker, after making allowance for corrected calibration. Spontaneous fluctuations of the pupil iron out well enough in averaging that consistent conditioning would not be concealed. While an experimenter may be subject to all sorts of errors of perception, and an objective record would be desirable, it does not seem possible to offer a satisfactory explanation of Baker's results on the basis of expectations and anticipations of the experimenter. Unfortunately, the arrangement he used in which he was both presenting stimuli and following the pupil maximizes the possibility of some influence by these factors. The experiment of Metzner and Baker (**6**), supporting Baker's original findings, is not a very exacting control on subjective factors, since the recording experimenter was informed by telephone of the beginning of a trial, although there was some uncertainty about the exact moment of onset of the sound. As Wedell, Taylor, and Skolnick (**10**) reported, quasi-conditioned responses may appear as a result of the subject's expectations (as well as the experimenter's) but these may be corrected for by appropriate control experiments.

5. Time intervals. Our stimulus intervals, and those of Wedell, Taylor, and Skolnick, duplicated those of Baker, so that this cannot be the difference between the experiments.

Because of our failure to reproduce his findings with respect either to auditory thresholds or to conditioning, we must conclude that Baker's conditions are not stated with sufficient completeness to delineate the circumstances under which his results can be reproduced.

Summary

Three experiments were designed to reproduce Baker's findings with respect to the conditioning of pupillary responses to near-threshold stimuli. An initial obstacle was encountered in the inability to secure a series of auditory threshold values in which percentages of confidence and percentages of success agreed with his. Adopting percentage of success as defining the intensity of sounds to be used, the first experiment consisted of attempted conditioned constriction (11 subjects) and dilation (10 subjects) to a tone of 1900 cycles. Results were negative. The second experiment was similar to the first, but binocular fixation at a nearer point (15 cm) was substituted for the fixation at 100 cm in Experiment 1. Results were again negative with attempted dilation (11 subjects) and at-

tempted constriction (10 subjects). In the third experiment, a noise was substituted for tone, and conditioned constriction attempted at three intensities for each of 19 subjects, again with negative results. Thus in experiments with 61 subjects no confirmation was secured for Baker's results.

(Manuscript received January 21, 1941)

References

1. Baker, L. E., The influence of subliminal stimuli upon verbal behavior, *J. exp. Psychol.*, 1937, **20**, 84–100.
2. Baker, L. E., The pupillary response conditioned to subliminal auditory stimuli, *Psychol. Monogr.*, 1938, **50**, No. 223, pp. v + 32.
3. Cason, H., The conditioned pupillary reaction, *J. exp. Psychol.*, 1922, **5**, 108–146.
4. Hudgins, C. V., Conditioning and the voluntary control of the pupillary light reflex, *J. gen. Psychol.*, 1933, **8**, 3–51.
5. Kotliarevsky, L. I., The formation of pupillary conditioned reflexes and of a differentiation in response to both direct and verbal stimuli, *Ark. biol. Nauk.*, 1935, **39**, 477–489. (*Biol. Abstr.*, 1937, **11**, No. 13724.)
6. Metzner, C. A., and Baker, L. E., The pupillary response conditioned to subliminal auditory stimuli: a control experiment, *Psychol. Bull.*, 1939, **36**, 625.
7. Miller, J. G., Discrimination without awareness, *Amer. J. Psychol.*, 1939, **52**, 562–578.
8. Steckle, L. C., Two additional attempts to condition the pupillary reflex, *J. gen. Psychol.*, 1936, **15**, 369–377.
9. Steckle, L. C., and Renshaw, S., An investigation of the conditioned iridic reflex, *J. gen. Psychol.*, 1934, **11**, 3–23.
10. Wedell, C. H., Taylor, F. V., and Skolnick, A., An attempt to condition the pupillary response, *J. exp. Psychol.*, 1940, **27**, 517–531.
11. Zabarenko, R. N., *Conditioning of the pupillary light reflex*, Unpublished Master's Thesis, Univ. Pittsburgh, 1939.

EXTERNAL INHIBITION AND DISINHIBITION IN A CONDITIONED OPERANT RESPONSE [1, 2]

BY R. M. GAGNÉ

Connecticut College

The phenomena of external inhibition and disinhibition were first named and described by Pavlov (**8**) in studies on the formation of conditioned salivary reflexes. He and his collaborators noticed that an extra stimulus, applied while a conditioned reflex is in the process of formation, produces a considerable reduction in the magnitude of the conditioned response. When an extra stimulus is introduced after experimental extinction has taken place, an increase in strength of the conditioned response results. According to Pavlov, the same stimulus which produces disinhibition (or increase in response magnitude) also produces external inhibition (a decrease in response magnitude) when applied during acquisition (**8**, p. 67).

Disinhibition of the conditioned galvanic skin response has been studied by Switzer (**15**) and by Wenger (**17**). In the latter's investigation the same extra stimulus which produced external inhibition was used to produce disinhibition. Razran (**9**) obtained a correlation of +0.91 between the amount of disinhibition and the degree of extinction of conditioned salivary responses in human beings. Hunter (**6**) demonstrated the phenomenon in the extinguished running response of rats on an octagonal electric grill. Reynolds (**11**, **12**) showed that an extra stimulus applied at a choice point in a maze results in the disinhibition of the appropriate response. The phenomenon has been studied in connection with voluntary responses by Rexroad (**10**) and by Hunter (**7**).

A failure to obtain disinhibition in a conditioned operant response has been reported by Skinner (**13**). The response of pressing a lever was established in eighteen rats and extinction curves, which measure the rate of responding, were then obtained. Several different extra stimuli, such as tossing the animals in the air, pricking the tails with a needle, the sound of the food magazine, and a light, were introduced as extinction approached completion. In no one of the twenty-four cases, Skinner reports, did the extra stimulus produce a disinhibitory effect. In several cases when the rate had for some reason become

[1] This is one of a series of three articles presented as a thesis to the faculty of the Graduate School of Brown University in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Psychology, October, 1940.

[2] The author is indebted to Professor C. H. Graham of Brown University, under whose direction the present research was carried out.

unusually slow, the extra stimulus served to bring the curve up to its 'proper' position. No case was found in which the rate increased sufficiently to bring the curve above its 'envelop' (an extrapolation of the initial portion of the extinction curve).

In this connection it should be noted that Horns and Heron (**5**), employing the Skinner lever-pressing technique, report a long-lasting increase in rate of responding when electric shock is applied during extinction of the response. These results throw some doubt upon the generality of Skinner's experimental findings.

In a recent paper Graham and Gagné (**2**) have used the concept of inhibition in a description of conditioning. If this description is to have significance it should be possible to demonstrate the necessity for this variable. From this point of view an experiment on disinhibition assumes importance, for the demonstration of disinhibition may be expected to necessitate the postulation of inhibition.

In the present experiment I have investigated the effects of two different external stimuli, applied during acquisition and extinction, upon the strength of a conditioned response. The extra stimuli used are (1) a loud buzzer and (2) a light scratching at the starting box of the apparatus. With the second of these two stimuli (scratch) it has been possible to demonstrate disinhibition and external inhibition in a response of the type used by Graham and Gagné (**2**). With the buzzer a different result is obtained, and this is considered in its relation to conditioning theory.

Apparatus and Procedure

The apparatus employed in the present experiment is similar to one described in the previous paper (**2**). It consists of a narrow runway 1¼ inches wide and 3 feet long. (The runway previously was ¾ inch wide.) At one end of this runway is a platform on which is placed a starting box, and at the other end is a similar platform for a food box. The two boxes are of identical construction and are used interchangeably during the experiment as starting box and food box. They measure in their outside dimensions 6 inches wide, 6½ inches high, and 7¼ inches long. In front of each platform is a small door which may be quietly raised and lowered by the experimenter.

Fifty-two white rats with an average age of three months were used in the experiment. Eleven additional animals were discarded because they did not leave the starting box in five minutes or less, on the first trial of acquisition. None of the animals employed had been used previously for experimental work.

The general procedure followed was the same as that described previously (**2**). Each rat was given four short training sessions on four consecutive days. This time was spent in the exploration of an elevated runway (of the same width as that used in the experiment) without any reinforcement. The animals were fed for twenty minutes a day in the food box, where they learned to seize and eat food of the size employed in the experiment. On the final training day the time for eating was shortened to twelve minutes. The experimental food consisted of small pellets of dog chow cut to a size which weighed about 0.04 gram apiece. Fifteen of these were eaten by the rat in the course of an experiment.

Each experimental session was conducted 24 hours after the final training period. A session consisted of (1) fifteen equally spaced trials of acquisition and (2) the number of trials of extinction which were necessary to produce the disappearance of the response for a period of 3 minutes.

Ten seconds before the beginning of each trial the experimenter placed a pellet of food in the food box and then put the starting box in position. A trial was begun with the opening of the door in front of the starting box and the simultaneous starting of a stop-watch. The watch was stopped when the animal left the starting box and passed a mark four inches in front of the door, on his way to the food box. This interval was recorded as the latent period of response. A new trial was started at the constant interval of seventy-five seconds after the animal arrived in the food box. During this interval the rat finished eating and was then carried in the box from the food platform to the starting platform. The box was placed in the latter position ten seconds before the next trial was to begin.

Acquisition was continued for fifteen trials, and the first extinction trial followed the fifteenth acquisition trial at the usual seventy-five-second interval. During extinction the procedure was changed only in the respect that no food was placed in the food box.

Two different extra stimuli were employed to test for the phenomena of external inhibition and disinhibition. The first of these was a fairly loud buzzer attached to the under side of the starting platform in such a manner that the clapper struck the platform. The second stimulus was produced when the experimenter scratched lightly on the back of the starting box. This stimulation, which will be referred to as a 'scratch,' caused the rat to turn around and look towards the rear of the box ('exploratory behavior'). The buzzer was sounded for four seconds and turned off two seconds before the beginning of the next trial. The scratching lasted about four seconds on the average, but in all cases was continued until the rat turned around completely to face the back of the box. The trial was begun about two seconds after the occurrence of this response.

Five groups of animals were subjected to five different experimental procedures. Group 1 was a control group consisting of twelve animals in which acquisition and extinction were measured in the usual way and with the introduction of no new extraneous stimuli. The eight animals in Group 2 were exposed to the buzzer for four seconds before the first trial of acquisition and before the fifth trial of extinction. Group 3 was composed of ten animals which were subjected to the buzzer before the fourth acquisition trial and the fifth extinction trial. In Group 4, the influence of the scratch on the first trial of acquisition and the fifth trial of extinction was measured in twelve rats. Finally, the scratch was presented before Trial 4 of acquisition and Trial 5 of extinction in the case of ten rats of Group 5. The differences in the latent period, if they exist, should appear by this procedure between the twelve rats of Group 1 and (*a*) eight rats with buzzer on Trial 1 of acquisition; (*b*) twelve rats with scratch on Trial 1 of acquisition; (*c*) ten rats with buzzer on Trial 4 of acquisition; (*d*) ten rats with scratch on Trial 4 of acquisition; (*e*) eighteen rats with buzzer on Trial 5 of extinction; and (*f*) twenty-two rats with scratch on Trial 5 of extinction.

Results

The values of log latent period for all groups of animals in acquisition and extinction are presented in Table 1, together with the average deviation of each value. These values are the average logarithms of the latent period obtained for each trial in each of the five groups. They represent the log geometric average of the data on each trial.

The Effect of an Extra Stimulus on the First Trial of Acquisition.—Figure 1 shows the results obtained when the buzzer and the scratch are presented just preceding the start of the first acquisition trial in two different groups of animals. The experimental points for the control group are not shown, but the curve of acquisition for this group is a best-fitting line drawn by inspection through these points. In the control group, the value of log latent period for Trial 1 was found to be 1.78, representing a time of 61 seconds. The value of the logarithm of the latent period following the presentation of the

buzzer, measured in the eight rats of Group 1, is 1.81 (65 seconds). When the scratch was presented before the first acquisition trial, the average log latent period for the rats of Group 4 was 1.76 (58 seconds). The differences between the values for the three groups are not significant.

TABLE 1

THE AVERAGE LOGARITHM OF THE LATENT PERIOD OF THE RESPONSE ON EACH TRIAL OF ACQUISITION AND EXTINCTION FOR FIVE GROUPS OF RATS

Group 1 (12 rats) is a control group. In Group 2(8 rats) the buzzer was presented preceding the first trial of acquisition and the fifth of extinction. Group 3 (10 rats) received the buzzer stimulus before the first trial of acquisition and the fifth of extinction. In Group 4 (12 rats) the scratch was presented before the first trial of acquisition and the fifth of extinction. Group 5 received the scratch stimulus before the fourth trial of acquisition and the fifth of extinction.

Process	Trial	Log L.P. Group 1 (Control)	A.D.log P.	Log L.P. Group 2 (Buzzer on 1)	A.D.log P.	Log L.P. Group 3 (Buzzer on 4)	A.D.log P.	Log L.P. Group 4 (Scratch on 1)	A.D.log P.	Log L.P. Group 5 (Scratch on 4)	A.D.log P.
Acquisition	1	1.78	0.22	1.81	0.29	1.70	0.23	1.76	0.24	1.77	0.24
	2	1.25	.30	1.42	.24	1.38	.25	1.29	.27	1.42	.24
	3	1.06	.25	1.14	.30	1.04	.26	0.99	.25	1.11	.25
	4	0.96	.22	1.01	.22	1.41	.16	1.01	.23	1.23	.34
	5	0.90	.31	1.04	.28	1.01	.33	0.83	.33	0.86	.51
	6	0.85	.34	0.88	.28	0.96	.24	0.82	.37	0.96	.38
	7	0.75	.36	0.70	.37	0.74	.31	0.73	.39	0.64	.39
	8	0.62	.29	0.80	.21	0.81	.22	0.56	.38	0.70	.39
	9	0.68	.27	0.55	.31	0.55	.21	0.54	.29	0.65	.33
	10	0.55	.33	0.59	.38	0.52	.38	0.56	.33	0.63	.27
	11	0.53	.24	0.54	.41	0.59	.21	0.55	.27	0.50	.37
	12	0.36	.30	0.43	.35	0.46	.30	0.47	.37	0.43	.33
	13	0.42	.29	0.48	.36	0.43	.34	0.45	.33	0.40	.37
	14	0.39	.28	0.37	.29	0.32	.27	0.32	.32	0.36	.32
	15	0.37	.24	0.32	.31	0.33	.26	0.31	.24	0.36	.33
Extinction	1	0.40	0.29			0.41	0.32			0.35	0.31
	2	0.79	.33			0.67	.27			0.67	.37
	3	1.07	.20		Groups	1.02	.21		Groups	1.07	.25
	4	1.21	.21		2 and 3	1.15	.22		4 and 5	1.24	.27
	5	1.35	.18		[Buzzer	1.59	.12		[Scratch	0.94	.28
	6	1.36	.17		on 5] [3]	0.99	.24		on 5] [3]	1.29	.27
	7	1.40	.22			1.38	.22			1.43	.29
	8	1.39	.20			1.49	.20				

[3] Although these groups have been subject to different experimental procedures during acquisition, the similarity of results during extinction justifies the procedure of considering them together.

It will be noted that the close correspondence between the values obtained on the first trial is continued throughout the course of fifteen trials of acquisition. The curve which represents the results for the control group is also seen to be a best-fitting curve for the data of the other two groups.

We may conclude from these results that neither the buzzer nor the scratch had any effect upon the animals' behavior which might be interpreted as facilitative or inhibitory. Before the acquisition has begun, we should expect that all external stimuli, including the conditioned stimulus itself, would have little effect upon the particular response which is to be measured. A similar argument is used to justify the inference of the value of response magnitude at the be-

ginning of acquisition, for the purpose of fitting data to theory (1). The fact that fairly intense stimulation immediately preceding the first trial can have so little influence upon the latent period may be taken as an additional bit of evidence for this point of view. The results also indicate that the effects which these extra stimuli have on later trials of acquisition and extinction (to be described below) are functions of the underlying processes involved, and not the result of facilitating or depressive properties of these specific stimuli.

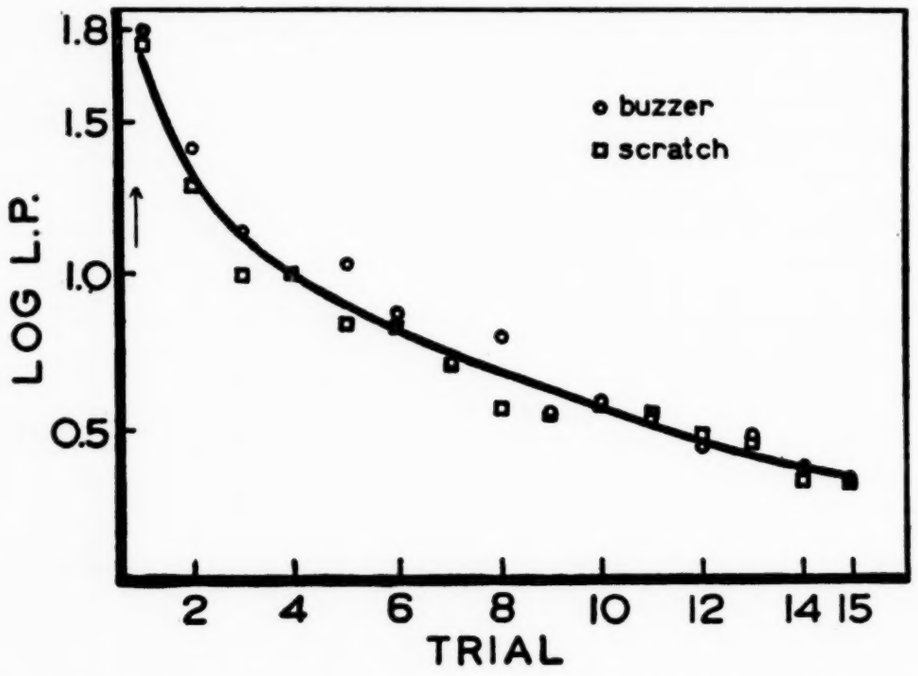

FIG. 1. The effect of the buzzer and the scratch on the first trial of acquisition. The arrow indicates the trial following the extra stimulus. The control group is represented by a smooth curve; experimental values are shown for the two groups in which the buzzer and the scratch, respectively, were presented before trial 1. There are no significant differences between the values for the three groups on any of the fifteen trials.

The Effect of an Extra Stimulus on the Response at the Fourth Acquisition Trial.—The results of applying the buzzer and the scratch preceding the fourth trial of acquisition, in two separate groups of rats, are shown in Fig. 2. The acquisition curve of the control group, represented by a smooth curve, has an experimental value of log latent period of 0.96 (9.2 seconds) at this point. The average logarithm of the latent period in the group which was subjected to the buzzer is 1.41 (26 seconds), and the difference of 0.45 log units is a highly significant one. The *t*-test of significance indicates that this difference has a probability of occurrence by chance of less than once in 100 times.

The animals in Group 5, which were stimulated by the scratch to turn around in the starting box, exhibit an average log latent period

of 1.23 (17 seconds) on the fourth trial of acquisition. The difference between this value and that of the control group is 0.27 log units. This difference is less than that produced by the buzzer, and the *t*-test of significance indicates a probability of chance occurrence of less than 5 percent. Although such a difference is slightly less significant than that effected by the buzzer there can be little doubt that its reality is strongly indicated.

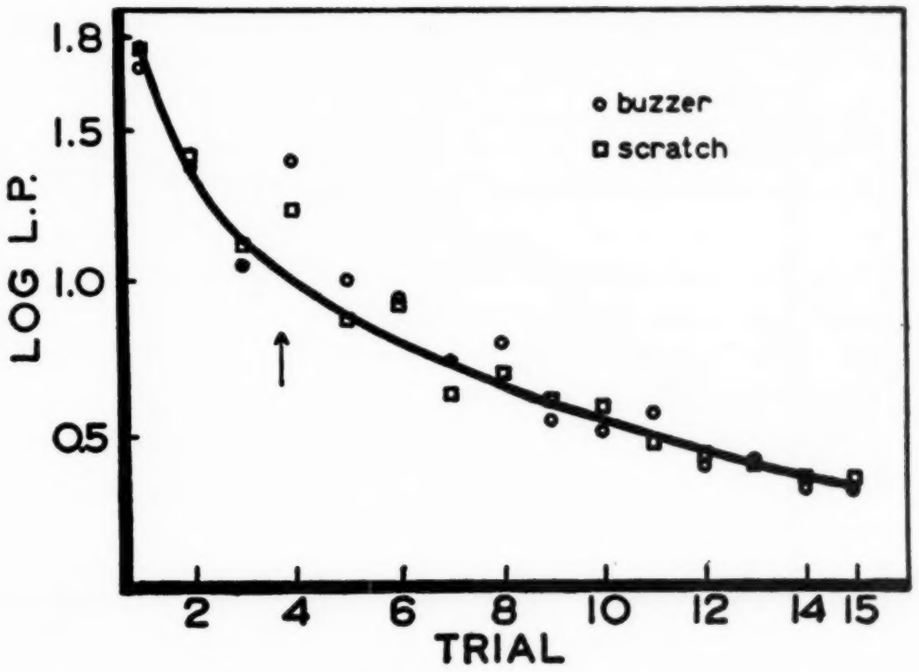

FIG. 2. The effect of the buzzer and the scratch on the fourth trial of acquisition. The control group is represented by a smooth curve, while experimental values are shown for the two groups in which the buzzer and the scratch, respectively, were presented just before Trial 4. On this trial the buzzer produced an increase of 0.45 log unit, and the scratch an increase of 0.27 log unit, over the experimental value for the control group. These differences are significant.

The results indicate that the distracting effect of the buzzer applied two seconds before the beginning of the fourth trial of acquisition causes a significant increase in the latent period of the response on that trial. This represents a decrease of strength of the conditioned response, and an instance of what has been called 'external inhibition.'

The introduction of the constant response of turning around may also be said to constitute an example of external inhibition since, when this response is elicited by scratching on the starting box just preceding the fourth trial, a considerable lengthening of the latent period results. It may be objected that this longer latent period might be simply the result of making it necessary for the animal to turn around towards the front of the starting box before making his response. However, observations of the animals' behavior on this

trial showed that the rats all turned around again almost immediately after the opening of the starting-door. The amount of time required for this activity could constitute only a small part of the total latent period, or of the difference in the latent periods of the two groups on this trial. Further corroboration of this point of view is found in the effect of the scratch during extinction. Here the turning around results in a decreased latent period.

The Effect of the Buzzer on the Response at the Fifth Extinction Trial.—The results represented in Fig. 3 show that the buzzer, when

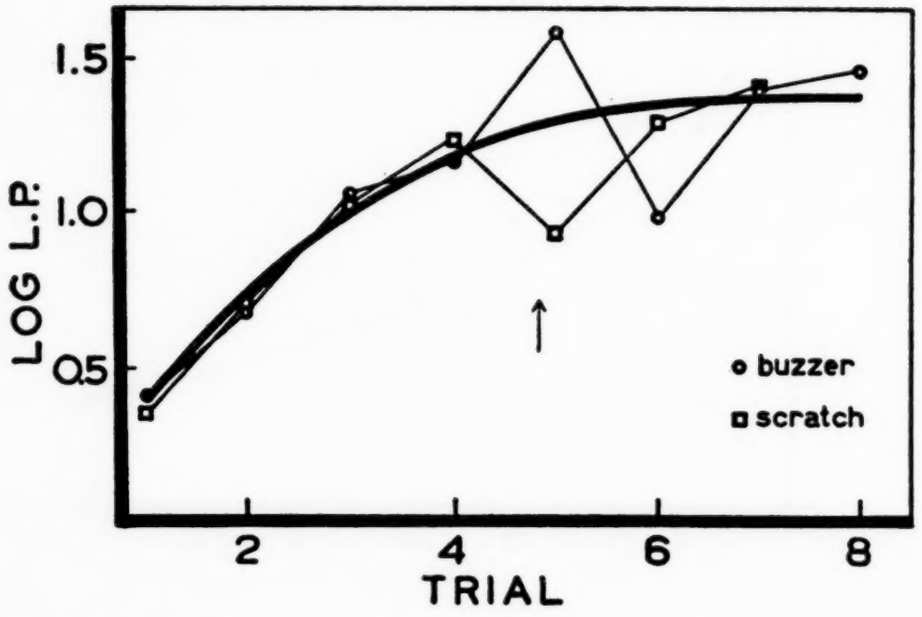

FIG. 3. The effect of the buzzer and of the scratch when each is applied preceding the fifth trial of extinction. Extinction in the control group is represented by a smooth curve, while for the other two groups the curves are constructed by joining experimental points. The buzzer produces (as compared with the control values) an increase of 0.24 log unit on the fifth trial, and a decrease of 0.37 log unit on the sixth. The scratch produces a reduction of 0.41 log unit in the latent period on the fifth trial. All of these differences are significant.

applied preceding the fifth trial of extinction, has the effect of increasing the latent period of the response on the fifth trial. The average logarithm of the latent period on this trial for the rats which were subjected to this procedure is 1.59 (39 seconds). This value is considerably above a control group log latent period value of 1.35 (22 seconds). The difference which occurs here is 0.24 log unit and one which, according to the *t*-test, might be expected to occur by chance in less than 1 percent of similarly conducted experiments.

These data reveal that the buzzer significantly reduces the strength of the conditioned response on the fifth trial of extinction, when it is presented two seconds before the beginning of that trial. This 'inhibitory' influence of the buzzer is also indicated by the observations of the animals' behavior. After the stimulus was presented, each of

the rats crouched slightly and maintained a relatively fixed position for as long as ten seconds in some cases. It may be tentatively inferred that the animals were to some extent 'frightened' by the buzzer. At any rate, the buzzer, by its effect upon the rats' behavior, resembles what Skinner (**14**) has called an 'emotional' stimulus.

The Effect of the Buzzer on the Response at the Sixth Extinction Trial.—Although the buzzer was presented only preceding the fifth trial of extinction, it was found to influence also the latent period of the response on the next following trial, a result exhibited in Fig. 3. The average logarithm of the latent period on the sixth extinction trial was found to be 0.99 (9.8 seconds). This is considerably lower than the control group value of 1.36 (23 seconds) for the same trial. The difference of 0.37 units is a highly significant one according to the *t*-test (probability of chance occurrence of less than 1 percent).

It may be concluded that the buzzer, after decreasing the magnitude of the response of the fifth extinction trial, has the effect of significantly increasing response strength on the sixth trial. The data thus exhibit a result which closely resembles the phenomenon of disinhibition. In this case a stimulus which acts to inhibit the response on both the fourth trial of acquisition and the fifth of extinction, apparently has a facilitative influence on the response at the sixth trial of extinction. The effect of this stimulus does not depend upon its inherent characteristics, but rather upon the state of the conditioned response which is measured.

This experiment does not determine whether the increase of the latent period on the sixth trial depends merely upon the passage of additional time after the presentation of the buzzer, or upon the occurrence of a lengthened latent period on the immediately preceding trial. An answer to this question, which would aid in the interpretation of the phenomenon, may be provided by future experiment.

The Effect of the Scratch on the Response at the Fifth Extinction Trial.—The experimenter scratched at the back of the starting box, causing the rat to turn and face the rear of the box, about two seconds before the fifth extinction trial was begun. The result of the introduction of this constant response on the latent period of the conditioned response on the fifth trial is shown in Table 1 and Fig. 3. The control group has an average log latent period on this trial of 1.33 (22 seconds), while the group subjected to the scratch has a mean log latent period of 0.94 (8.7 seconds). In this case the difference between the means of control and experimental groups is 0.41 log unit. The *t*-test indicates that the probability of chance occurrence of this difference is less than 1 percent.

The shortened latent period on the fifth trial of extinction, following shortly after the introduction of the extra response of facing the back of the starting box, is the measure of a considerably increased magnitude of response on this trial. Again our results constitute evidence for the reality of the phenomenon of disinhibition. Since the effect of the same stimulus applied during acquisition is to decrease the strength of response (as measured by a lengthened latent period) the scratch cannot reasonably be termed a facilitative stimulus.

Taken as a whole, the results of this investigation may be said to confirm the observations of Pavlov in regard to the phenomena of external inhibition and disinhibition. They also constitute further evidence of similarity between the operant and respondent types of conditioned responses (*cf.* **18**). The data show, first of all, that neither of the two stimuli used has any significant part in determining the magnitude of the conditioned response on the first trial of acquisition. Secondly, both the buzzer and the scratch produce external inhibition when presented before the fourth trial of acquisition. Thirdly, the buzzer has an inhibitory effect even when applied during extinction just preceding the fifth trial. On the next following trial, the sixth of extinction, disinhibition is evidenced by an increased magnitude of response (decreased latent period) which appears without the application of any further extraneous stimulation. Finally, the scratch, which has an inhibitory effect during acquisition, causes an increase in response strength at the fifth extinction trial.

Discussion

This investigation has shown that an extra stimulus introduced at different stages of the formation or extinction of a conditioned operant response can have entirely opposite effects upon the magnitude of that response. It may, as in the case of the scratch, depress the response during acquisition, and increase it during extinction. Or it may, like the buzzer, have a depressing effect whenever applied, but an effect which is followed by an increase of response magnitude during extinction.

Skinner's (**14**) proposal is that conditioning results in the formation of a 'reserve' which is depleted in extinction. A single process therefore underlies the measured strength of response during each of these events. An extra stimulus may, according to this author, have a facilitative effect, or an 'emotional' effect, in which case it depresses the rate of responding by arousing conflicting emotional responses. In the situation employed in the present experiment, the buzzer may for more than one reason be called an 'emotion-producing' stimulus.

First, its effect on the animals' behavior is greatly to reduce activity and to cause the momentary assumption of a fixed, crouching position. Second, this stimulus consistently produces a decreased strength of the immediately following conditioned response. This evidence may be taken as supporting Skinner's position with regard to disinhibition. An increased rate of responding in the lever-pressing response, following an emotionally depressed rate, was the only evidence Skinner (**13**) found for 'disinhibition.' In our experiment an emotion-producing stimulus effects a depression of response strength which is followed on the next trial by an increase in strength.

A scratching at the back of the starting box preceding the first and fourth trials of acquisition and the fifth of extinction causes the animal to make the response of turning around. This extra response has no effect on the first acquisition trial, but causes a depression of the response on the fifth extinction trial. These results are more difficult to explain in terms of Skinner's hypothesis. A stimulus which facilitates the response during extinction should have, according to this idea, a similar effect during acquisition. Although Skinner (**13**) failed to obtain disinhibition with several different extra stimuli, none of these clearly resembled the scratch in its production of a constant extra response.

An acceptance of the idea of inhibition makes possible a reasonable interpretation of the facts which the present experiment has revealed. Such an interpretation is more closely related to the proposals of Pavlov (**8**, p. 67) than to those of some other theorists.

In this experiment the clearest evidence concerning inhibition is found in the effects of the 'scratch' on response strength during acquisition and extinction. Since these effects are opposite in the two situations, it is not possible to relate their occurrence to any property of the stimulus. The difference in effect must rather be attributed to differences in the underlying determinants of response strength. Using Pavlov's terms we may say that the 'scratch' stimulus interferes with the dominant excitatory process which is being increased as the conditioned response is established, thereby revealing an underlying inhibition. During extinction, the inhibitory process is the dominant one, and the interference provided by the scratch decreases the effect, as shown by an increased response magnitude.

It may be pointed out that the results obtained with the scratch stimulus are in line with the interference hypothesis of Guthrie (**3**) and Wendt (**16**). This is especially true because the 'scratch'clearly produces a conflicting response when it is applied during acquisition and during extinction. The difference between the Guthrie-Wendt and the Pavlovian hypotheses seems to be one chiefly of emphasis. Both agree in accounting for external inhibition as the result of an

increase in strength of conflicting responses (**8**, p. 44). Disinhibition is interpreted by Pavlov as the 'inhibition of inhibition,' and similarly, according to interference notions, as the external inhibition of conflicting responses (**4**, p. 117). When one deals experimentally with modifications of some particular response, it is more immediately practicable to measure increments and decrements in that response than to attempt the measurement of all possible conflicting responses. Conceptually, there is much to be gained by representing 'interference' as an inhibitory process which tends to diminish the strength of the measured response.

The effect of the buzzer on response strength during acquisition and extinction may also be accounted for in terms of the concept of inhibition. It may be supposed that the buzzer stimulus acts to increase the amount of 'internal' inhibition in such a way as to decrease the strength of response whenever it is applied. This is revealed as a reduction in response magnitude on the trial following the application of the buzzer, which occurs during both acquisition and extinction.

As this extra inhibition dies away, spontaneous recovery may be thought of as occurring to produce an increased magnitude of response on the subsequent trial. During acquisition the amount of spontaneous recovery may be slight relative to the amount of the dominant excitation process which is the chief determinant of response strength. Consequently the added spontaneous recovery is not sufficient to produce an appreciable increase in response magnitude. During extinction, however, the situation is different in that inhibition is now the dominant process. Since, according to Pavlovian principles, the amount of spontaneous recovery is proportional to the degree of inhibition, we should expect the recovery process to have a considerable effect in augmenting response strength. In the present experiment this effect is shown clearly as an increased magnitude of response on the sixth extinction trial, following the application of the buzzer on the fifth.

Because of its consistent depressing effect on the response during both acquisition and extinction, the buzzer may be called 'emotion producing' as well as 'inhibitory.' This interpretation is consistent with the ideas of Skinner (**14**). Whether the difference between the effects of the two extra stimuli used in the present experiment may be attributed to differences in stimulus intensity, or to some other factor, remains a problem for future experimentation.

In summary, the effects which are produced by the extra stimuli employed in this experiment may be interpreted in several possible ways. (1) The effect of the buzzer in producing a depression of response strength whenever applied, may be interpreted according to

Skinner's hypothesis as an 'emotional' effect. In this event the increase of response strength which occurs during extinction on the trial following the depression, may be considered a result of the compensatory increase in the number of available responses. (2) The buzzer may be conceived in the light of Pavlov's interpretation as an 'inhibitory' stimulus, which first acts to depress the strength of response by increasing the amount of (internal) inhibition. Following this effect, as inhibition dies away an increase in spontaneous recovery may produce an increase in response magnitude on the next trial. (3) An extra stimulus, such as the 'scratch,' may arouse an extra response which conflicts with the entire reaction system of the conditioned response. Such an effect is similar to what Pavlov called 'external inhibition,' producing a depression of conditioned response magnitude during acquisition and an increase (disinhibition) during extinction. In this case the 'inhibition' is entirely consistent with the ideas of Guthrie and Wendt, since it is conceived as the interference of one total reaction system with another. Contrary to Skinner's suggestion, the disinhibitory effect of this stimulus cannot reasonably be attributed to facilitation.

Taken together, the results of the present experiment demonstrate the necessity for the use of the concept of inhibition in a theoretical formulation of the conditioning process. The most important evidence bearing on this point is the occurrence of disinhibition during extinction, after the application of a clearly non-facilitative stimulus.

Summary

1. This experiment has been concerned with the question of the influence of extra stimuli during acquisition and extinction of a conditioned operant response. The response, measured in five groups of white rats, was one which required the animal to traverse a simple elevated runway from a starting box at one end to a food box at the other end. The time taken for the rat to leave the starting box was recorded as the 'latent period' of the response. The extra stimuli were (1) a buzzer and (2) a scratch at the back of the starting box. The buzzer was applied for four seconds and discontinued two seconds before the beginning of the next trial. The scratch was continued in each case until the animal turned around to face the back of the starting box.

2. Neither the buzzer nor the scratch had an appreciable effect upon the latent period of the response when applied just before the first trial of acquisition.

3. The buzzer produced an increased latent period of response when applied before the fourth trial of acquisition. When applied

before the fifth trial of extinction it resulted in an increase in latent period on that trial. This increase was followed on the sixth trial of extinction by a decrease in latent period. These effects were shown to be significant.

4. The scratch produced (*a*) a significantly increased latent period of response on the fourth trial of acquisition, and (*b*) a significantly decreased latent period of response on the fifth extinction trial.

5. The data provide evidence for the occurrence of external inhibition and disinhibition in a conditioned operant response. It is shown that the disinhibition produced cannot be attributed to facilitation.

(Manuscript received January 23, 1941)

References

1. Gagné, R. M., and Graham, C. H., The effect of an 'emotional state' on the initial stages of acquisition in a conditioned operant response, *Proc. nat. Acad. Sci.*, Wash., 1940, **26**, 297–300.
2. Graham, C. H., and Gagné, R. M., The acquisition, extinction, and spontaneous recovery of a conditioned operant response, *J. exp. Psychol.*, 1940, **26**, 251–280.
3. Guthrie, E. R., *The psychology of learning*, New York: Harper, 1935, pp. viii + 258.
4. Hilgard, E. R., and Marquis, D. G., *Conditioning and learning*, New York: Appleton-Century, 1940, pp. xi, 429.
5. Horns, H. L., and Heron, W. T., A study of disinhibition in the white rat, *J. comp. Psychol.*, 1940, **30**, 97–102.
6. Hunter, W. S., The disinhibition of experimental extinction in the white rat, *Science*, 1935, **81**, 77–78.
7. Hunter, W. S., An experiment on the disinhibition of voluntary responses, *J. exp. Psychol.*, 1938, **22**, 419–428.
8. Pavlov, I. P., *Conditioned reflexes: an investigation of the psysiological activity of the cerebral cortex* (Trans. by G. V. Anrep), London: Oxford Univ. Press, 1927, pp. xv + 430.
9. Razran, G. H. S., Decremental and incremental effects of distracting stimuli upon the salivary CRs of 24 adult human subjects (inhibition and disinhibition?), *J. exp. Psychol.*, 1939, **24**, 647–652.
10. Rexroad, C. N., Reaction time and conditioning: extinction, recovery, and disinhibition, *J. exp. Psychol.*, 1937, **20**, 468–476.
11. Reynolds, H. E., The disinhibiting effect of an electric shock upon the maze performance of the white rat, *J. comp. Psychol.*, 1936, **22**, 187–197.
12. Reynolds, H. E., Further disinhibition phenomena in the maze behavior of the white rat, *J. comp. Psychol.*, 1939, **27**, 271–282.
13. Skinner, B. F., A failure to obtain 'disinhibition,' *J. gen. Psychol.*, 1936, **14**, 127–135.
14. Skinner, B. F., *The behavior of organisms: an experimental analysis*, New York: Appleton-Century, 1938, pp. ix + 457.
15. Switzer, St., C. A., Disinhibition of the conditioned galvanic skin response, *J. gen. Psychol.*, 1933, **9**, 77–100.
16. Wendt, G. R., An interpretation of inhibition of conditioned reflexes as competition between reaction systems, *Psychol. Rev.*, 1936, **43**, 258–281.
17. Wenger, M. A., External inhibition and disinhibition produced by duplicate stimuli, *Amer. J. Psychol.*, 1936, **48**, 446–456.
18. Youtz, R. E. P., Reinforcement, extinction, and spontaneous recovery in a non-Pavlovian reaction, *J. exp. Psychol.*, 1938, **22**, 305–318.

ELECTROENCEPHALOGRAPHIC INDIVIDUAL DIFFERENCES AND THEIR CONSTANCY: I. DURING SLEEP [1, 2]

BY CHARLES E. HENRY [3]

Brush Foundation, Western Reserve University School of Medicine

In the application of the electroencephalographic technique to problems of psychology, two major lines of approach appear evident. One of these involves the study of intra-individual variability in the EEG as such variability may be related to moment-to-moment variability in psychological process. The other involves the study of inter-individual variability in the EEG as such variability (maintenance of constant EEG differences between individuals) may be related to differences in the psychological makeup of different individuals. Data are in existence which may be brought to bear on both of these problems, but relatively little has been done toward clarifying the relationships between them. The importance of these relationships for the evaluation of the EEG as a technique in psychology has been emphasized by Knott (**15**),* who has pointed to the fact that inter-individual differences imply a measure of the EEG which is constant and which does not, due to its nature, take into account the characteristics of intra-individual variability. Since, however, both types of measurement are made on the same series of phenomena, one must know, before generalizing from one set of measurements, what the relationships between them are. The fact that long-time, or supposedly constant, characteristics of the EEG have been described (Davis and Davis, **2**; Travis and Gottlober, **23**, **24**; Rubin, **22**; Williams, **27**) must somehow be fitted with the fact that more variable, short-time characteristics have been described and correlated with short-time psychological variables (Travis, **25**; Travis and Egan, **26**; Hoagland, **11**; Knott, **13**; Hadley, **8**, **9**).

Two logically independent assumptions might be made to fit the concepts of long-time constancy and short-time variability and to

[1] The main content of the first part of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, in the Department of Psychology in the Graduate College of the State University of Iowa. August, 1940.

[2] This study is one in a series dealing with the electroencephalographic correlates of psychological processes, under the direction of Dr. John R. Knott in the laboratories of theDepartment of Psychology in the State University of Iowa.

[3] The author wishes gratefully to acknowledge his indebtedness to Dr. John R. Knott, who directed this research.

* The references will be given at the end of a second article which will appear in the next issue of this JOURNAL.

link them together. The first would state that individual differences in the EEG are determined genetically and are relatively unalterable; intra-individual variability then occurs within dimensions thus established. The second assumption would state that individual differences in the EEG are determined by environmental (historical) factors and are alterable; intra-individual variability then occurs within dimensions thus established.

The tests and evaluations of these assumptions depend to a considerable extent upon the electrocortical characteristics subjected to measurement. Ever since Gibbs (**5**) pointed out that the EEG must be viewed as a continuum of frequencies, the tendency to measure isolated segments of the frequency continuum has been under attack. However, the fact that measurement of certain aspects of the alpha frequencies (from 8–13 cycles per second) has become a well-established habit among workers in the area has led many to erect hypotheses solely on the basis of such measurements. Other frequencies, known to represent abnormal cortical functioning, may yield data which would be contraindicative to those secured by measurement of alpha. As an example, Gottlober (**6**) found no tendency for measurements of alpha to be significantly related within family groups; Lennox, Gibbs and Gibbs (**17**) have presented data indicating that the slow frequencies associated with epilepsy are hereditary. Due to the disparity of the general conclusions regarding heredity, per se, of the two investigations, one must be somewhat cautious in setting up experimental approaches to the assumptions outlined above.

Perhaps a third assumption, more compatible with present-day knowledge, would lead to more productive experimental enterprise. This assumption would state that certain characteristics of individual differences are genetically determined, that certain characteristics of individual differences are environmentally determined, and that these factors, acting together in a manner not yet known, limit intra-individual variability.

Since the alpha characteristics are the most easily measured variables in the *normal* EEG, and since for the majority of psychophysiological correlative work the normal is the type of behavior sampled, alpha may be sampled with reasonable assurance that a not too-great myopy is being employed. The tests of the third assumption, however, may not be able to measure hereditary factors by sampling alpha alone.

Knott, Henry and Hadley (**14**) have recently attempted to evaluate the first two assumptions by means of observations upon individuals widely differing in the amount of alpha rhythm (in time) which they exhibited. They reasoned as follows: If individual differences in the EEG which are defined in terms of the percentage of

time an alpha rhythm is present on the record (alpha index) are, as has been suggested by Davis (1), evidence of genetically determined differences in cortical organization, then differences between individuals in conditions other than the standard one in which the alpha index is measured, should be found. They believed that measurements of other-than-alpha frequencies should correlate with alpha measurements if the Davis hypothesis were true.

Selecting individuals on the basis of extreme mean differences in percent time alpha during the waking state, they found that the records of their subjects regressed to a comparable level of electrocortical activity during sleep, the condition chosen as other-than-standard. Mean differences in the percentage of time various sleep rhythms were present did appear between the high and the low alpha groups, but these were overshadowed by variability within the groups.

They did find some correlation apparent between waking alpha index and the indices of the sleep rhythms. This finding seemed at the time to assume importance in the light of the hypothesis of Jasper and Andrews (12) that the slow, random waves of sleep are derived from the waking alpha rhythm, and that the spindle frequencies are derived from the waking beta rhythm. Since those individuals with little waking alpha ought then to show little derived alpha during sleep (*i.e.*, random waves), the data incontrovertibly indicate that low alpha subjects are potentially high alpha, since they showed large amounts of random.

If this test of the Jasper-Andrews hypothesis is valid (and it may not be, as will be seen somewhat later), the hypothesis that the percent time waking alpha is gentically determined would not be supported with great vigor, unless some factor responsible for the repression of alpha during waking were the genetic variable. This leads rapidly to cycle and epicycle and might be better discarded at the outset.

Problem

It appears necessary to extend the observations of Knott, *et al.*, in order that a more thorough examination of the assumptions may be made. They used extremes of waking alpha, and although this served for preliminary work, the equivocality of the data make it essential to draw a sample of subjects distributed throughout the continuum of alpha indices and to study them under several experimental conditions.

Thus, a sample of continuously varying alpha indices was selected, and it was studied under three general levels of activity: the standard waking condition of minimal activity; the condition of sleep; and the condition of psychological activity, which was subdivided into various

degrees or types.[4] The standard condition was conceived to be a 'zero' or reference point, and the other conditions to be variations below and above that point.

These conditions were selected to cover their wide range in order that sufficient variation would be likely in the EEG; this variation could then be evaluated with reference to the zero level chosen.

Apparatus and Procedure

A four-channel amplifier system, coupled to four Grass moving-coil, ink-writing oscillographs was used in obtaining the records. Each channel consisted of a four-stage push-pull resistance-capacitance coupled amplifier which actuated a bridge-type impedance-matching stage used to drive the ink-writer. A time constant of .67 seconds was used. The frequency response of each channel was uniform up to 100 cycles per second. A single section *m*-derived high-pass filter ($m = .6$) was usually inserted in one of the ink-writing circuits. This filter sharply reduced in amplitude all frequencies below ten cycles per second and made it possible to record lower voltage spindle waves during higher voltage slow random patterns. A signal pen recorded the presentation of a 180-cycle tone used in the experiments to determine whether or not the subject being studied was asleep. 'Sleep' was defined as having begun when this tone, at moderate intensity, failed to elicit a response.

Recording tape speeds of 15, 30, and 60 mm per second were available, the slowest speed being the one most commonly used in recording the sleep records. At this speed, eight hours of record could be recorded on about fourteen hundred feet of tape.

Surface electrodes (flattened solder discs, 3–4 mm in diameter) were used. They were attached with collodion; resistances before sleep were always less than 5000 ohms, and were usually below 15,000 ohms the next morning. The majority of the sleep records were secured with monopolar leads on the left occipital, left motor and left frontal areas. Supplementary records from the left sensory area, as well as records from the vertex and contralateral homologous areas, were obtained occasionally for exploratory and confirmatory purposes. These data are not discussed in this report. The ground electrodes were of the same disc type and were attached in parallel to the lobe of each ear. A selector switch, controlled outside the sleeping room, made it possible to place eleven leads, each of which could be connected to any amplifier for any combination of either bipolar or monopolar recording.

The subjects slept on a comfortable cot in an electrostatically shielded, semi-soundproofed and well-ventilated room. The selector switch and ventilating equipment was in a room enclosing the sleeping room, with the amplifiers and recording equipment in still another adjacent room. Constant verbal contact between subject and experimenter, and selective contact between experimenter and subject was provided by an intercommunication system. A microphone suspended directly over the subject made possible the detection of movement, respiration, talking in sleep, etc.

Twenty-two all-night sleep records were obtained from twenty male subjects between 20 and 45 years of age, one subject being run on three consecutive nights. The subjects were selected so that five individuals fell in each of the quarters of the alpha index range (1–25, 26–50, 51–75, 76–100) as designated by Davis and Davis (2). These are referred to as rare alpha, mixed alpha, subdominant alpha and dominant alpha, respectively.

Analysis of Data.—Each minute of record from the time of retiring until after the subject awakened in the morning was read with respect to the predominant type of rhythm according to the following categories:

Day record, with varying degrees of alpha rhythm.

Sleep alpha, the occasional bursts of ten per second waves found when the subject is believed to be asleep.

Low voltage, the record characteristically found as the individual becomes drowsy and begins to drop off to sleep. It is composed of arhythmic waves, usually below 25 μv. in amplitude.

Spindles, the rhythmic bursts of approximately 12 to 14 per second waves, each burst lasting from one-half to several seconds, ranging up to 50 μv.

[4] The data from this third level of activity are to be presented separately in a later paper.

Spindles plus random, waves of from 50 to 300 μv., at varying frequencies from one-half to six per second. The spindles are superimposed upon or fall between the random waves.

Random, frequencies and voltages as above, but with spindles so inconspicuous as to be non-discernible even on the filtered channel.

These categories were scored in terms of percent of time spent in sleep.

To evaluate the differences between the groups of subjects, analysis of variance and the *t*-test were employed (18).

Results

The P.M. waking alpha indices, computed from at least three meters of record, formed a roughly graded series extending from 90.4 to 1.7 percent. The means of the four groups all fell within 3 points of the mid-point for their respective quarter of the alpha index scale. Table 1 presents the waking and sleep records in terms of group (dominant, etc., alpha) means for the three cortical areas recorded. Figure 1 graphically presents the group means for all conditions studied.

The analysis of variance indicated that in terms of P.M. waking alpha index the subjects formed a heterogeneous distribution. The *F*-ratio (Table 4) was significant at the one percent level.

Sleep alpha was usually found in conjunction with movement on the part of the subject, though at times it appeared with no detectable

TABLE 1

Mean Percentage of Time Each Wave Type Appeared in the Left Occipital Motor and Frontal Area Sleep Records of the Dominant, Subdominant, Mixed and Rare Alpha Groups

	P.M. Waking Alpha	Sleep Alpha	Low Voltage	Spindles	Spindles +Random	Random	A.M. Waking Alpha
Left Occiput							
Dominant alpha.....	84.7	3.82	18.92	32.72	44.23	0.32	65.2
Subdominant alpha.....	65.8	2.67	17.69	49.63	27.56	2.45	63.5
Mixed alpha...	39.2	2.43	19.53	45.84	31.66	0.54	47.2
Rare alpha....	9.6	0.69	19.06	35.55	44.10	0.58	29.5
Left Motor or Vertex							
Dominant alpha.....		2.70	20.27	32.94	44.09	0.00	
Subdominant alpha.....		2.24	16.18	43.82	35.81	1.96	
Mixed alpha...		2.43	19.55	45.16	32.39	0.47	
Rare alpha....		0.69	18.96	33.73	46.31	0.29	
Left Frontal							
Dominant alpha.....		2.70	20.27	33.66	43.37	0.00	
Subdominant alpha.....		2.24	15.85	47.72	29.57	4.60	
Mixed alpha...		2.43	19.55	45.21	32.34	0.47	
Rare alpha....		0.70	17.32	37.83	43.95	0.18	

accompanying movement. (The appearance of muscle action potentials on the record formed a quite sensitive indicator of movement, appearing at times when no noise could be detected through the microphone suspended less than a foot above the shoulders of the subject.) It was arbitrarily decided that *sleep alpha* must be present for a minimum duration of one second in order to be included in this classification; otherwise the sleep pattern preceding and following was tallied as having no change. Table 1 shows that the dominant

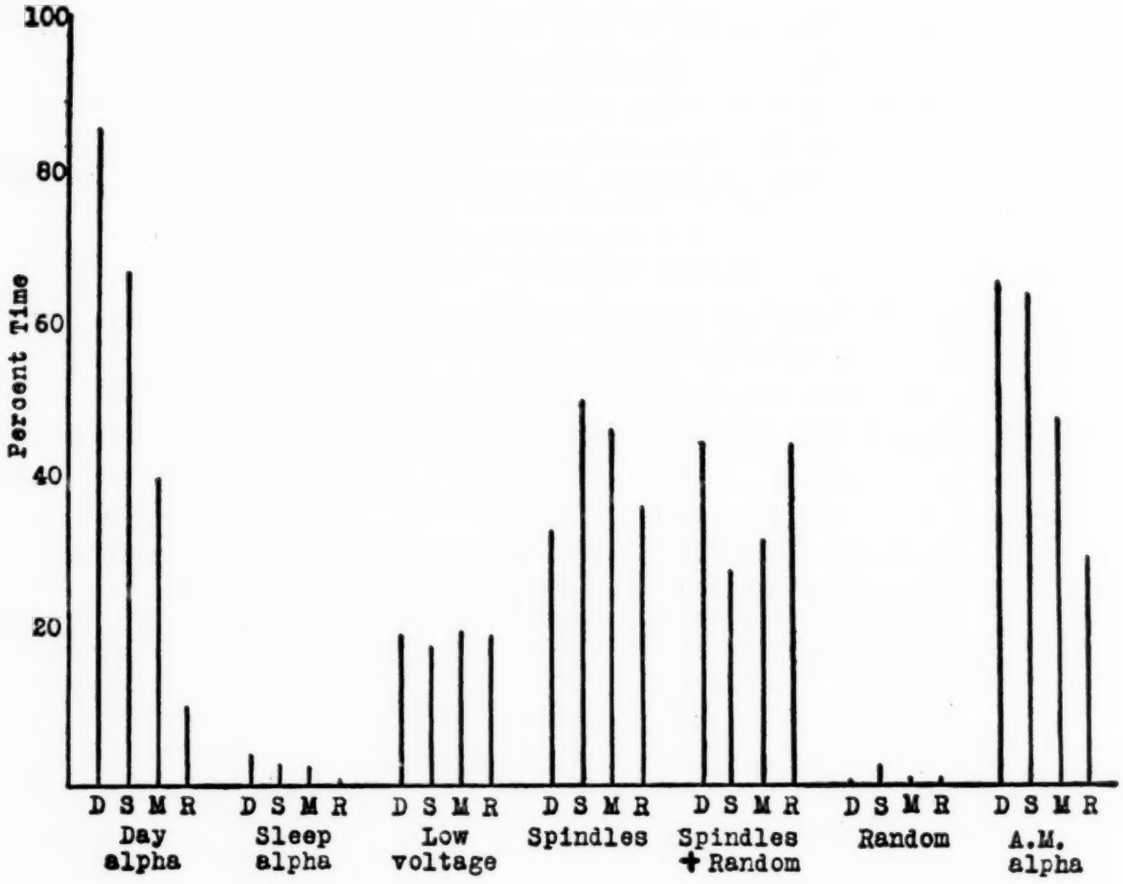

FIG. 1. Mean values of P.M. alpha, the various sleep rhythms, and A.M. alpha for the four groups of subjects (dominant, subdominant, mixed, and rare alpha), as selected according to P.M. alpha indices.

alpha group tended to have the greatest amount of *sleep alpha* and the rare alpha group to have the least. The difference between the groups was 3.13 percent as compared to 75.1 percent for the waking condition; this difference in sleep alpha was not significant (Table 3). The correlation between sleep alpha and P.M. alpha was .66 ± .09 (*cf.* Table 2).

The differences in the mean amount of *low voltage* shown throughout the night were less than two percent for the left occipital area, and slightly more than this for the other areas. There were no consistent trends, the individual ranges within each group being very

comparable. The correlation with P.M. alpha was .05. Short periods of this pattern were the general rule, although towards morning longer periods of *low voltage* tended to become more frequent. Due to these small differences, *F*-ratios were not computed.

Group mean differences in the amount of *spindle* activity were more marked.[5] The lack of linear trend, however, was shown by the fact that both the dominant and rare alpha subjects appeared to have less of this pattern than did the subdominant and mixed alpha sub-

TABLE 2

RANK ORDER CORRELATIONS BETWEEN ALPHA INDEX AND PERCENT TIME SLEEP RHYTHMS (*LO*, *LM* and *LF* areas, in that order) *

	Sleep Alpha	Low Voltage	Spindles	Spindles +Random	Random	A.M. Alpha Index
P.M. alpha index	.66±.09	.05±.16	.00±.16	−.12±.16	−.24±.15	.58±.11 *LO*
Sleep alpha		.12±.16	−.27±.15	−.04±.16	−.09±.16	.42±.13 *LO*
		.13±.16	−.11±.16	−.21±.16	.05±.16	*LM*
		.10±.16	−.37±.14	−.01±.16	.22±.16	*LF*
Low voltage			.03±.16	−.46±.12	−.10±.16	−.36±.14 *LO*
			−.07±.16	−.50±.12	−.20±.15	*LM*
			.05±.16	−.59±.11	−.31±.15	*LF*
Spindles				−.79±.06	.05±.16	−.21±.16 *LO*
				−.75±.07	−.07±.16	*LM*
				−.74±.07	−.01±.16	*LF*
Spindles +random					.37±.14	.27±.15 *LO*
					.45±.13	*LM*
					.36±.14	*LF*
Random						.10±.16 *LO*

* $N = 20$ except for A.M. correlations, in which $N = 19$.

jects. The left occipital data appeared to be heterogeneous by the analysis of variance. The *F*-ratio indicated significance at the five percent level. Application of the *t*-test indicated that only two groups, the dominant and the subdominant, differed significantly (five percent level). The correlation between day alpha index and *spindle* activity was zero.

The *spindles plus random* patterns also showed more variability, the trend here being the reverse of that for *spindle* activity. The dominant and rare alpha groups showed the greater amounts of this

[5] *Spindles* has been accepted as a condition of true sleep by all investigators, to which the present study found but one possible exception. One subject showed well defined bursts of *spindles* at a time when he was 'certain' he was awake, but drowsy. Until further evidence is available, a more conservative interpretation of this datum would be that the subject, in the drowsy state, was unable to distinguish clearly between the waking and the sleeping states. It is well known that such a condition (drowsiness) is not favorable for accurate introspective reports.

pattern, the subdominant and mixed, the lesser amounts. The *F*-ratios did not indicate any of the differences between groups to be significant. The correlation between day alpha index and *spindles plus random* was −.12.

Mean differences in the amount of *random* activity are probably not important, since but nine of the twenty subjects showed this pattern unaccompanied by spindles [6] and these were not distributed equally in the groups. In general there appeared to be a tendency for more of the subjects with a low waking alpha to show at least some of this pattern. The correlation with P.M. alpha was −.24.

Mean differences in the amount of alpha activity recorded in the morning (after sleep) are of importance. A period of approximately five minutes elapsed between awakening and the recording of the A.M. alpha index, during which time the light was on, the experimenter entered the room and talked with the subject, and the subject stretched about in order to become thoroughly awake. In the subsequent recording, stimuli were presented at once if the record showed any indication that the subject was becoming drowsy. It will be noted from Table I and Fig. I that the A.M. means of all four of the groups approached the mid-point (50 percent) of the alpha index range. The rare and dominant alpha group means changed more than did the mixed and subdominant alpha group means. These changes in alpha index were as follows:

	P.M.	A.M.	Change
Dominant alpha	84.7	65.2	−19.5
Subdominant alpha	65.8	63.5	− 2.3
Mixed alpha	39.2	47.2	+ 8.0
Rare alpha	9.6	29.5	+19.9

The magnitude of the mean change for the dominant and rare alpha groups was nearly identical, but opposite in direction. The magnitudes of the mean change for the subdominant and mixed alpha groups were not as near equality, but the directions were again reversed. Six of the nine subjects with day alpha indices over 50 percent shifted downward in the morning; eight of the ten subjects with day alpha indices of less than 50 percent shifted upward.[7]

[6] In an earlier report (14) we found little or no *spindles plus random* activity, for which there were three mutually contributing reasons. First, we were recording largely from the occipital area, where this pattern is less distinct; second, we did not have the benefit of specially tuned filters to emphasize waves of this frequency; third, if amplification is kept at or below the point where the high voltage *random* is accurately recorded, the lower voltage *spindle* waves are inconspicuous. The *random*$_2$ which we reported at that time has now been recorded on several other subjects in both the waking and sleeping states. We are, until the possibility is specifically ruled out, terming this an artifact due to an excess of sweat on the scalp near the electrode, since our more recent '*random*$_2$' has been associated with just such a condition.

[7] This discrepancy in number of subjects arises from the fact that it was impossible to obtain a record after sleep from one of the subjects in the subdominant alpha group.

The importance of the non-significant *F*-ratios pertaining to the sleep data becomes more evident in the light of an analysis of variance for the A.M. alpha indices (Table 4; the magnitude of the ratio necessary for significance is the same as in Table 3).

It has already been seen that the subjects were heterogeneously distributed prior to sleep. After sleep, however, the intra-group

TABLE 3

F-RATIOS FOR SLEEP DATA

Area	Sleep Alpha	Low Voltage	Spindles	Spindles + Random
Left occiput	1.97	—	3.32	2.28
Left motor	1.83	—	1.99	1.24
Left frontal	1.31	—	2.18	1.43

1 percent level 5.29
5 percent level 3.24

variance reached a magnitude comparable to that of the inter-group variance, and the subjects formed a homogeneous rather than a heterogeneous distribution. In other words, the *same subjects* formed a more homogeneous sample not only during sleep but upon awakening in the morning, than they did before sleep. This indicates a reduction of individual differences.

TABLE 4

MEAN DIFFERENCES AND *F*-RATIOS BEFORE AND AFTER SLEEP

Group	Before Sleep		After Sleep	
	Difference	1%	Difference	1%
Dom. *vs.* subdom.	16.5	14.4	1.7	29.5
Subdom. *vs.* mixed	29.0	14.4	16.3	29.5
Mixed *vs.* rare	29.6	13.6	17.7	26.1

P.M. alpha, $F = 100.34$
A.M. alpha, $F = 1.34$

This conclusion is further validated by mere inspection of the group means. Prior to sleep, the mean of each group was significantly different from those groups adjacent to it as well as those groups removed from it. After sleep, however, there was a wide degree of overlap between adjacent groups; the dominant alpha group, as a matter of fact, did not differ significantly from the remote mixed alpha group (Table 4).

The data on the subject who was recorded for three consecutive all-night periods are presented in the following tabulation. Since the

first (and only) night's sleep was used for each of the other subjects, only the first night's sleep for this individual was used in the group means. Only the data from the left occipital area are presented.

Day Alpha	Sleep Alpha	Low Voltage	Spindles	Spindles + Random	Random	A.M. Alpha
66.8	6.10	15.37	49.48	29.06	0.00	30.6
80.0	6.39	15.86	21.51	55.31	0.92	58.4
71.5	2.82	14.99	37.79	44.40	0.00	41.2

The negative relationship between P.M. alpha index and *spindles*, and the positive relationship between P.M. alpha index and both *spindles plus random* and A.M. alpha index is difficult to evaluate. The day alpha index varied over a range of 13 percent; still other records taken from this subject under 'standard' conditions yielded alpha indices as low as 57 percent, although on the whole he was classified in the subdominant group. The effect of including either the second or third night's data (or both) in the subdominant group would be that of making this group even more comparable to the other groups.

The data pertaining to the left motor and left frontal areas were similar to those for the left occipital area. Eight of the twenty subjects (2 dominant, 1 subdominant, 3 mixed and 2 rare alphas) showed such similar patterns from the three areas that it was impossible to distinguish categorically between them. When differences were present, they were usually of the order of five percent or less, with a correspondingly small effect on the group means. This, however, does not indicate perfect similarity or synchrony between the several areas during sleep. The classification in general use for sleep potentials is sufficiently broad to permit some slight differences to be present and yet not be described as a distinct wave category. *Spindles*, for example, tended to be most clearly observed from the motor area or vertex, yet they were almost always present on the other areas at the same time; but the bursts were shorter and less distinct. The same statement may be made with regard to synchrony; occasionally what appeared to be perfect synchrony was observed between several areas, although it was more usual for there to be only the same *general pattern* present, with slightly different frequency and amplitude characteristics for each area, hence asynchronous phase relations. Research directed more specifically toward such differences has been presented by other workers (**3**, **19**, **20**), and has not been a prime concern of this report.

Throughout the course of these experiments, dreams were occasionally reported. Six of the subjects reported a total of eight dreams,

usually upon awakening spontaneously during the night. The pattern preceding the report of dreaming varied between *low voltage*, *spindles*, and *spindles plus random*, and in addition has shown in every case one or more clearly defined bursts of alpha waves. These bursts have all occurred within one minute of the time that the dreams were reported. Rather than constituting evidence that dreams and alpha waves are related, these data seem to indicate that

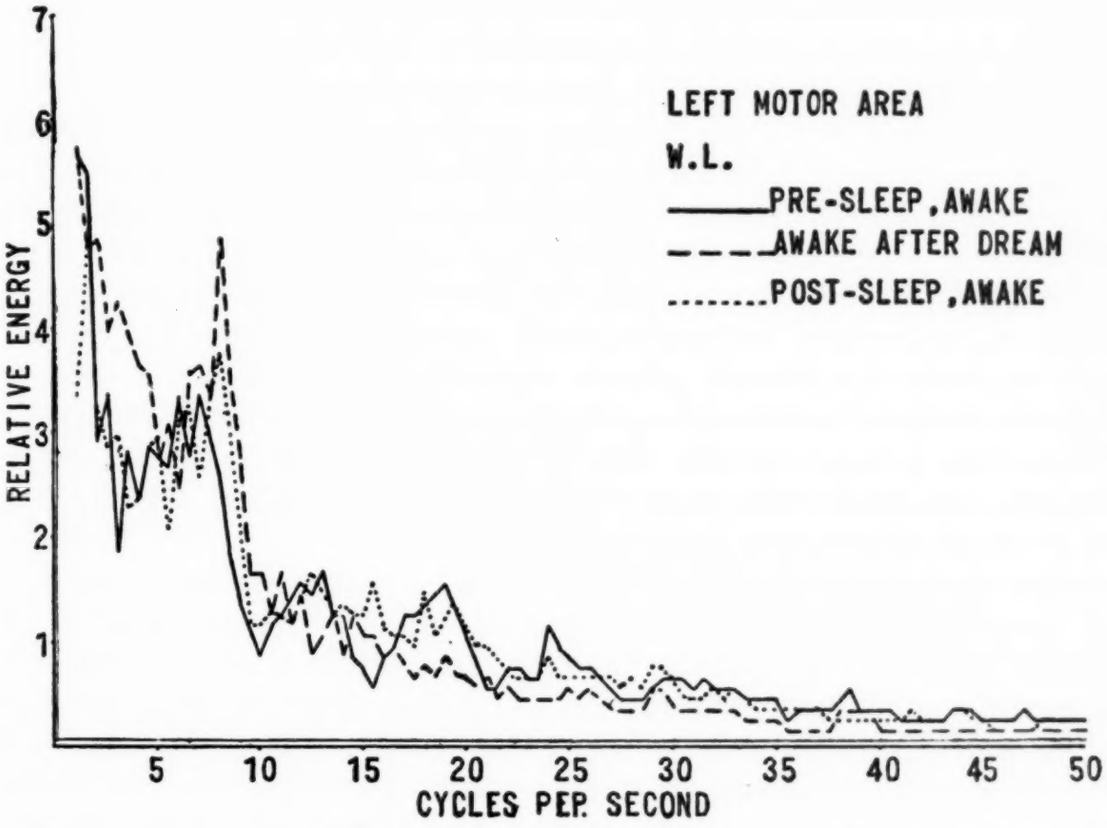

FIG. 2. Fourier transforms of three EEG records secured on rare alpha subject W. L. The disparity between the energy values in the alpha band in the pre-sleep awake and the post-sleep awake is clear. (These records were analyzed by Dr. F. A. Gibbs, Boston City Hospital, by means of the Grass low-frequency wave analyzer.)

dreams may be related to at least transient shifts upward in the level of sleep.

As previously pointed out, there was a tendency for individuals with low P.M. alpha indices to show an increased alpha index upon awakening in the morning. Figure 2 presents a more complete account of these changes than may be gained from the numerical designation of percent time alpha alone. The pre-sleep alpha index of this subject amounted to 7.2 percent, and the post-sleep (awake) alpha index has increased to 17.6 percent. The corresponding increment in energy at the alpha band is clearly evident in the transforms. Even more interesting, however, was the amount of alpha activity exhibited upon awakening during the night after a dream.

Although he was not awake long enough at this time to compute an alpha index on three meters of record, the transform data certainly corroborate the protocol notation that he showed 'lots of alphas' at this time.[8]

Discussion

These data may be evaluated from two interlocking viewpoints, one from that of the EEGs secured during sleep, the other from that of the EEGs secured shortly after awakening from sleep.

The data from the sleep records have extended and corrected the somewhat tentative results previously reported by Knott, *et al.* Whereas they found, on the basis of a smaller number of subjects which included only the extremes of the alpha index continuum, that some quantitative differences might be expected to exist between the sleep records of these two groups, and that these differences might be related to the initial grouping in terms of alpha index, the present observations failed to demonstrate significant differences between continuously distributed groups of subjects, and likewise failed to find correlations between sleep rhythms and the waking alpha rhythm. Thus, the present results clearly indicate a regression of electroencephalographic differences during sleep, a regression which Knott, *et al.*, suggested rather than demonstrated.

On the basis of the Jasper-Andrews hypothesis that random waves of sleep are derived from the waking alpha rhythm, these results would not be predicted, for the dominant alpha group should show, according to the hypothesis, more random activity than the rare alpha group, and the subdominant and mixed alpha groups should be systematically distributed between these extremes. On the surface this may appear to be a rejection of the hypothesis on the basis of its predictive failure.

However, this technique of verification is dependent upon a basic assumption which has not been subjected to rigorous experimental scrutiny until the present time; namely, the assumption that the alpha index is a comparatively stable measure of a fixed, unalterable capacity of a given individual cerebral cortex to produce an alpha rhythm. The assumption must, therefore, be substantiated before the test of hypothesis can be regarded as relevant.

The original alpha indices prior to sleep were distributed throughout the range of alpha indices, and the four groups of subjects studied were discrete by empirical selection. Analysis of variance demonstrated the groups to be heterogeneous, as would be expected. However, in the morning immediately after awakening, the clean-cut

[8] This is in good agreement with earlier observations reported by Knott, Henry and Hadley to the effect that four of their five rare alpha subjects showed a 'good' alpha rhythm at some time or other during the night.

differentiation of the four groups was lost. Whereas the P.M. groups were significantly different one from another, the same groups (of subjects) were not significantly different one from another except in the case of the more remote groups. Analysis of variance demonstrated the total group to be homogeneous. Emphasis should be placed on the fact that this was not the result of a loss by the originally high groups, but by both a loss of the high groups and a gain of the low groups. Therefore, one can apparently conclude that the alpha index is not a stable and unalterable phenomenon; and if it is not, then it cannot represent a fixed unalterable capacity for the production of the alpha rhythm.

The 'test' of the Jasper-Andrews hypothesis is thus not a test, and some other method of evaluating it must be found.[9]

This shift in alpha index clearly indicates the necessity of denying the *wholly* genetic view of the determination of the characteristic of the EEG which it measures. This lack of a previously asserted constancy makes some discussion of the factors involved in the alterability of the alpha index essential. The specific factors cannot as yet be described or named, although Freeman (4) has recently attempted to conclude that alpha index is related to the general excitation level of the organism. His causal arguments are open to debate. The following remarks are essentially tentative and general, and they are merely offered as suggestive of further research.

It may be that the general level of tension or reactivity of the organism is more uniform at this particular time in the wake-sleep cycle. It seems reasonable to suppose that the usual activities of the individuals and their general habits of reacting to the world at large had not yet been fully called into play. In short, it may be suggested that the behavioral patterns ('personality') which serve to differentiate individuals were not fully functioning at the time of day these samples were taken. (How would one rate 'personality' on the basis of behavior at this time?) This suggestion would be supported by the fact that worry, apprehension, attention, etc. seem to influence the percent time alpha of an individual. The result is usually a decrement, so that the assumed later increments for approximately half of the subjects would not fit this schema too neatly.[10]

[9] A different approach has been attempted by Knott, Henry and Gibbs (unpublished) with results strongly suggesting the plausibility of the Jasper hypothesis.

[10] Dr. Hallowell Davis, in a personal communication to the author, has suggested that two opposing tendencies in the determination of the alpha index may be operating during this early-morning recording. One of these is closely linked with drowsiness, which reduces alpha. The other of these is what Davis calls 'the alpha optimum' tendency, which increases alpha. In the middle range of alpha indices these would be expected to balance each other; at the upper extreme, the 'alpha optimum' tendency could find but little further expression, while drowsiness might operate to reduce the alpha index; at the lower extreme, however, drowsiness could reduce alpha

It would be surprising if there were no behavioral traits related in some way to the amount of alpha rhythm or to some other (perhaps more total and integrative) characteristic of the EEG. Lemere (16) has presented entirely qualitative evidence that cyclothymic persons have a 'better' alpha rhythm than schizothymic persons; and Gottlober (7) has attempted to relate the extroversion-introversion dimension of personality in normal subjects to percent time alpha. The latter failed to demonstrate adequately the relationship he stated he had discovered, but a reanalysis of his data by Henry and Knott (10) indicates some marginal relationship to hold. No data regarding specific, operationally defined items of personality and their relationship to the EEG have as yet appeared. Further work is clearly indicated in this area.

The point to be made with respect to these correlative attempts is that personality is too likely to be a learned series of responses for investigators to ignore the possibility (and perhaps the probability) that personality may determine certain EEG characteristics rather than be determined by the processes which they supposedly represent.

At the present time, although we cannot state that the 'long-time' characteristics of the EEG, such as the alpha index, are environmentally determined, we can state that they are not entirely determined by fixed (and perhaps genetic) variables. The 'history' of the individual has, in the present observations, been seen to change such 'long-time' characteristics. How this historical factor is involved in the determination of moment-to-moment variability is a problem which has yet to be attacked, but the issues seem to have become somewhat clearer on the basis of the delimitation of some determiners of inter-individual variability.

Summary and Conclusions

1. It has been suggested by previous workers that the alpha index of the human EEG is a relatively fixed and stable measure which is characteristic for a given individual, and subsidiary to this it has been suggested that this relatively fixed and stable measure may be indicative of an inborn cortical organization.

but little, while the 'alpha optimum' tendency might be encountered sufficiently often to produce a mean increase. The approximate equality of the mean alpha index shift for the high and low alpha groups, together with the statistical reliability of the data, make it most unlikely that these changes were the result of chance variability. The tendencies hypothecated by Davis could, of course, 'explain' the observed changes, although theoretical economy would tend to lead to the assumption that the alpha index is not constant, but depends on the conditions at the time of recording. Systematic exploration of conditions at the time of recording can then lead to lawful statements of the determining factors which underlie the alpha index.

2. It has been demonstrated that fluctuations in the alpha rhythm which are moment-to-moment in character are related to moment-to-moment fluctuations in psychological process.

3. These two types of EEG differences, inter-individual and intra-individual, must be related for they are the result of different types of measurement applied to the same data.

4. Three hypotheses are suggested to express the relationship of inter- and intra-individual differences: (1) Individual differences in the EEG are determined genetically and are relatively unalterable; intra-individual variability occurs within dimensions thus established. (2) Individual differences in the EEG are determined by environmental (historical) factors and are alterable; intra-individual variability occurs within dimensions thus established. (3) Certain characteristics of the EEG are genetically determined; other characteristics are environmentally determined; these factors, acting together in a manner not yet known, determine the limits of intra-individual variability.

5. The third hypothesis is the one most compatible with existing data for earlier statements concerning genetic determinants of the EEG dealt only with the alpha characteristics, and these represent only one series in a continuum of frequencies; this band has not been found to be determined by hereditary factors, while other bands may be.

6. If the alpha rhythm recorded in the standard condition does represent a fixed inborn cortical organization, this organization should be made manifest in other indices of electrocortical activity in other conditions.

7. The relation of waking alpha rhythms to rhythms recorded during sleep has been investigated by observations of the EEG in all-night sleep records of twenty adult subjects.

8. This population, heterogeneous with respect to waking alpha index prior to sleep, was homogeneous with respect to all sleep rhythms.

9. Waking alpha index *after sleep* indicated the population groups to be homogeneous and not heterogeneous as before sleep.

10. These findings indicate that individual differences in alpha during waking are not predictive of individual differences in other characteristics during other conditions, and that individual differences in alpha during one waking state are not predictive of individual differences in alpha during another waking state.

11. Thus, it may be concluded that environmental (historical) factors condition the production of alpha, and that evaluation of alpha in terms of inborn differences (or evaluation of inborn differ-

ences in terms of alpha) which are postulated without further evidence to represent inborn cortical differences must be qualified.

12. Relative to the relation of inter-individual differences and intra-individual differences in alpha, both are at least in part a function of environmental factors.

(Manuscript received January 21, 1941)

THE EFFECTS OF MONOCULAR VISION ON MEASURES OF READING EFFICIENCY AND PERCEPTUAL SPAN *

BY CHARLES ANTHONY KNEHR

Introduction

Since the first photographic recording of eye movements by Dodge and Dearborn at the turn of the century, a wealth of studies has been made on various phases of the reading process by means of this technique. Vernon [1, 2] and Tinker [3, 4] have extensively reviewed the various photographic techniques and much of the work which has been done to the present. In addition to studies related primarily to reading, as such, has appeared the use of the technique of eye movement recording for determining ocular behavior in such activities as typing or reading music, and the effects of drugs and oxygen lack on the use of the eyes.

Concomitant with the increasing interest in the experimental investigation of reading, and the use of measures of oculomotor behavior as they relate to activities other than reading, has appeared the increased use of eye movement photography in schools and educational clinics for the practical purpose of diagnosing poor readers and for routine prognosis similar to achievement tests. These workers are dependent upon preceding research for the correct interpretation of data obtained in the form of eye movement photographs.

The most familiar measures of ocular behavior found in a photographic record of reading are the succession of alternating pauses (fixations) and the movement of the eyes to a new pause or fixation along a line of printed matter. These movements (called saccadic) between the pauses are very rapid, requiring roughly seven percent of the total time for reading. The remainder of the time is occupied by the fixations. If the technique of obtaining and measuring the

* This dissertation was submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Faculty of Philosophy, Columbia University. The study was carried out under the guidance of Prof. A. T. Poffenberger with helpful advice from Prof. R. S. Woodworth and Dr. Conrad Berens. The author wishes to acknowledge his appreciation to them and to thank Dr. R. A. McFarland for the use of apparatus. Marion Knehr kindly aided in the preparation of the manuscript. The study was carried out in the Fatigue Laboratory, Harvard University, Boston, Massachusetts.

[1] Vernon, M. D. *The experimental study of reading.* Cambridge: Cambridge Univ. Press. 1931.

[2] Vernon, M. D. *The movements of the eyes in reading. Spec. Rep. Ser. med. Res. Coun., Lond.*, 1930, No. 148. Pp. 45.

[3] Tinker, M. A. Eye movements in reading. *J. educ. Res.*, 1936, 30, 1–37.

[4] Tinker, M. A. Experimental study of reading. *Psychol. Bull.*, 1934, 31, 98–110.

records is carefully used and proper precautions taken, the reading time and the number of fixations (fixation frequency) as described above may be used as valid and reliable measures of reading efficiency.[5]

Since the usual photographic eye movement record exhibits two traces, one for each eye, and can be assumed to be a fair approximation of ocular behavior at least for the horizontal, it would appear that the extent of cooperation of the two eyes would manifest itself in the record. Upon close inspection of the record, we do find a measurable variation in the distance between the traces for the two eyes indicating a variation in the amount of convergence during reading. The usual eye movement photograph shows a fairly large widening of the distance between the traces at the beginning of a fixation, usually more marked for the first fixation on the line, with a subsequent narrowing of the distance during the remainder of the fixation. Due to reversion by the lenses of the double eye movement camera, the movement of the visual axes relative to one another is opposite to what is measured in the photographic record. Consequently, the increase in distance between the traces at the beginning of a fixation is indicative that the visual axes have increased their convergence, while the subsequent narrowing of the distance between the traces apparently represents a relaxation of convergence. These variations, which we shall refer to hereafter as vergence movements, are assumed to be a measure of the precise coordinate activity of the two eyes.

There is an obvious lack of agreement, however, on the relation to reading efficiency of coordination or its lack, and on the effects of muscular anomalies and other known ocular defects. Specifically the disagreement touches on the magnitude of vergence movements and the evidence of rhythm in the photographic record. In addition the limitations of perceptual span in relation to reading efficiency are not clear. Eventually, the argument falls back into a discussion of whether peripheral or central factors are more important determiners of reading efficiency. The peripheral factors would include coordination and general oculomotor behavior, ocular defects and the area of clear vision, in distinction from what we label as perception, and familiarity with words, ideas, and meanings.

Concerning the role of these factors upon which there appears to be lack of agreement, let us turn first to a statement by Luckiesh and Moss.[6]

[5] Tinker, M. A. Reliability and validity of eye-movement measures of reading. *J. exp. Psychol.*, 1936, **19**, 732–746.

[6] Luckiesh, M., and Moss, F. K. *The science of seeing.* New York: D. Van Nostrand Co., 1937. P. 470.

> "The efficiency of the visual mechanism in reading depends very largely upon the character of the eye movements in reading." . . . "Since each of the many movements which occur during the reading of a single line involves the coördination of the activity of the six extrinsic muscles of each eye and an extremely close coördination of the movements of both eyes as well as minor changes in other ocular muscles, *it is obvious that even slight muscular anomalies impose severe handicaps in reading* (italics mine)."

This strong statement is not supported by much of the evidence to follow. A shade more cautious, but in the same general vein, is the following statement which appears in a pamphlet that accompanies a commercial eye movement camera.

> Using the technique that the Ophthalmograph and Metronoscope now make available to the profession, research workers have discovered that *causes* (italics mine) of poor reading ability include defective vision, poor muscle coördination, and lack of rhythm in eye movements. It was also found that the *inefficient reader has a narrower span of recognition* (italics mine) than the efficient reader—he is unable to visually perceive as many words at a glance.

This pamphlet recommends that the eye movement record may be used as a check on the muscle balance of the eyes.[7] The assumption must therefore be that phorias (measures of muscle balance of the eyes) are closely related to ocular behavior (especially vergence movements) as recorded photographically.

Clark,[8] working with muscular anomalies and their relation to vergence movements, found that patients with exophoria (tendency for the eyes to turn outward) showed overconvergence at the beginning of a line of print to a greater extent than other subjects, though the difference was not significant. This finding would appear to lend support to the preceding contention that phorias may be detected by means of eye movement photographs. Patients with a decidedly high degree of exophoria (15–24 diopters) showed significant differences as compared to a control group. But, Clark adds,

> the extent of binocular imbalance was *far from perfectly related* (italics mine) to the extent of the overconvergence movements, *e.g.*, the subject with the largest binocular imbalance (twenty-four diopters) made corrective movements which averaged sixty-eight minutes while the subject with the smallest binocular imbalance (fifteen diopters) made corrective movements which averaged one hundred forty minutes.

Other evidence on the relation of these peripheral factors to reading efficiency points in the opposite direction. McFarland, Knehr, and Berens [9] found little or no correlation between the extent of coöperation of the two eyes as measured by vergence changes and the phorias (muscle balance) both for a control group and for a series of clinical patients. Neither was there any apparent correlation

[7] Anonymous. Rhythm reading. *The Ophthalmograph.* American Optical Co., 1935

[8] Clark, B. Additional data on binocular inbalance and reading. *J. educ. Psychol.*, 1936, **27**, 473–475.

[9] McFarland, R. A., Knehr, C. A., and Berens, C. The effect of anoxemia on ocular movements while reading. *Amer. J. Ophthal.*, 1937, **20**, 1204–1221.

between other findings of the ophthalmological examination and the direct measures of reading efficiency. Similar findings are reported by Imus et al.[10] In a preliminary study on several hundred students they found that when subjects are grouped according to the diagnosis of ocular defects, there were no significant differences among them in *gains* or *performance* on reading tests, eye movement records, or academic points. Stromberg [11] finds almost identical distribution of esophoria and exophoria among fast and slow readers. However, fast readers consistently show higher ductions (or muscle strength of the eyes). In another study [12] the same investigator finds that acuity and emmetropia (with similarity of refractive condition of the two eyes) are of little consequence so far as reading speed is concerned and suggests that we must look to central factors for the final answer.

Stromberg [13] finds no difference between fast and slow readers in respect to vergence movements (co-ordination) in the eye records. Sisson [14] found with twenty-nine poor readers that eye movement training was no more efficacious, as regards reading rate, than was reading without intent to improve.

On the basis of the above studies there is ground for considerable doubt as to the role of peripheral factors such as coördination and span in reading efficiency. As if in summary we find this statement by Tinker:[15]

> It is probable that, in practically all cases, faulty eye movements are merely sumptoms rather than causes of poor reading.

This lack of agreement in the interpretation of coördinate activity of the two eyes as evidenced in eye movement photographs, and of ocular defects and span in relation to reading efficiency, furnished the background for the present investigation. The immediate stimulus for the study was the following statement by Mills:[16]

> Persistent attempts to do the impossible task of enlarging the span of visual accuracy by flash and speed training (in reading) not rarely lead to potential and actual deviations of the eyes.

---

[10] Imus, R. A., Rothney, J. W. M., and Bear, R. M. An evaluation of visual factors in reading. *Dartmouth College Publ.*, 1938.

[11] Stromberg, E. L. The relation of lateral muscle balance and the ductions to reading speed. *Amer. J. Optom.*, 1937, **14**, 415ff.

[12] Stromberg, E. L. The relationship of measures of visual acuity and ametropia to reading speed. *J. appl. Psychol.*, 1938, **22**, 70–78.

[13] Stromberg, E. L. Binocular movements of the eyes in reading. *J. gen. Psychol.*, 1938, **18**, 349–355.

[14] Sisson, E. D. Eye movement training as a means of improving reading ability. *J. educ. Res.*, 1938, **32**, 35 ff.

[15] Tinker, M. A. The use and limitations of eye-movement measures of reading. *Psychol. Rev.*, 1933, **40**, 381–387.

[16] Mills, L. The functioning of the eyes in the acquisition of an education. *J. Amer. med. Ass.*, 1929, **93**, 841–845.

The statement was evidently derived from clinical experience rather than from experimental investigation, and yet, because of its bearing on the previous discussion, it was very suggestive, both as regards the problems of the present investigation and as suggesting an experimental approach.

If we assume that the statement of Mills has a basis in fact, the implication is that attempts to speed reading result in the use of a single eye, with some tendency to deviation of the other eye. The question immediately arises, *why* reversion to moncular vision should result from attempts to read rapidly. The reason might be that perfect binocular fusion takes time, and that dispensing with fusion would leave the eyes free to advance more rapidly along the line of print. If however it should be found that monocular reading does not increase the speed and efficiency of reading, then the ocular deviations would require some other explanation. Mills (*op. cit.*) suggests that a reversion to the use of one eye may result in part from an abnormal use of peripheral vision. But since peripheral (even paracentral) vision is limited by a large decrease from foveal acuity, the suggestion is not particularly promising.

For the same reason, attempts to enlarge the perceptual span would appear to invite failure, so far as the span is dependent on peripheral factors, acuity in particular, for there is no likelihood of improving retinal acuity by any course of training. Any success in enlarging the perceptual span must depend on central factors such as increasing the familiarity of words and meanings.

These considerations led to the present investigation. The aim was to determine the role of monocular and binocular vision in reading, as shown by the usual measures of reading efficiency for short lengths of material, with especial reference to the role of perceptual span and coördinate behavior of the eyes.

More specifically the study was designed to test the following suggested hypotheses:

1. By dispensing with the need for binocular fusion of the two eyes (by the use of monocular vision) reading efficiency for short lengths of material, as measured by reading time and fixation frequency, should be materially increased.
2. Reading efficiency is intimately related to the perceptual span as a limiting factor. A small span should require a greater fixation frequency.

In addition, new evidence on the interpretation of vergence movements in reading appeared probable, by a comparison of changes observed in passing from binocular to monocular reading.

PROCEDURE

Subjects.—Fifteen men and one woman (subject Het) served as subjects for the experiment. Fourteen were students in the Graduate School of Business Administration at Harvard and two were staff members of the University. Ages ranged from twenty-one to thirty-eight with the mean at twenty-five. The main criterion (other than availability) for selection of subjects was that they be relatively free from visual defects (Table 1). Two of the sixteen subjects wore recently obtained corrections for refractive errors which are at the extreme of the so-called normal range, and which represent the order or refractive error that separates correction from non-correction in usual practice. On the whole the subjects may be considered as being visually within the normal range.

TABLE 1

SUMMARY OF DATA RELATING TO ACUITY, NEAR POINT, MUSCLE BALANCE AND OCULAR DOMINANCE

Name	Age	Acuity			Near Point of Convergence	Near Point of Accommodation		Lateral Imbalance			Ocular Dominance *			Rx
		OU	OD	OS		OD	OS	E	O	X	Right	Ambi.	Left	
Mos..	28	20/15−1	20/15−2	20/15	120/300	130/300	130/300		x				x	(OD+.50−.75 cyl. ax. 90) (OS +.50−.50 cyl. ax. 75)
Kar..	22	20/20+4	20/20	20/15−1	80/300	90/300	100/300	1			x			None
Alt...	21	20/15−1	20/20	20/20−1	170/300	200/300	150/300		x		x			None
Rak..	21	20/15+4	20/20+3	20/10−1	90/300	110/300	130/300			3			x	None
Fle...	22	20/20	20/20−3	20/20−1	130/300	110/300	110/300		x		x			None
Sew..	23	20/15	20/15−2	20/15−1	80/300	80/300	80/300			3	x			None
Wil...	26	20/20	20/20−1	20/20−1	120/300	140/300	130/300		x		x			+.50 ou
Goo..	21	20/20−2	20/20−2	20/25−2	120/300	110/300	110/300	2			x			None
Pet...	22	20/10	20/15	20/15−2	100/300	110/300	120/300			4	x			None
Dud..	33	20/20	20/20−3	20/20−2	170/300	160/300	130/300	2			x			None
Tay..	26	20/15−1	20/15−1	20/20−1	130/300	120/300	120/300			2	x			None
Kun..	25	20/20	20/20−1	20/20−1	110/300	100/300	100/300	2			x			None
Het..	38	20/20	20/20−1	20/20−1	180/300	170/300	170/300			4		x		None
Mor..	26	20/20+4	20/20	20/20−2	100/300	110 300	100/300			1	x			None
Rha..	25	20/20	20/20−1	20/20	100/300	110/300	110/300		x		x			None

* Peep-hole test.

Key of abbreviations

OU Both eyes.
OD Right eye.
OS Left eye.
E Esophoria—tendency for visual axes to turn inward at distance tested (40 cm).
X Exophoria—tendency for visual axes to turn outward at distance tested (40 cm).
O Orthophoria—no deviation of the eyes at distance tested.
Rx Prescription for lenses worn at time of test.

Eye Movements.—Each subject read four paragraphs before the eye movement camera[17] under the following conditions: The subject read a paragraph with both eyes, another with the right eye alone, one with the left eye alone, and a final paragraph alternately with the left and right eyes. More simply, the subject read with binocular vision, or with monocular vision (right, left, or alternating). In reading with the right or left eye alone, a card was held in front of the opposite eye. To obtain the effect of alternation of the separate eyes, a white bar, 30 mm wide, was placed between the eyes and the reading material at such a distance that the left eye saw the left half of the line, and right eye the right half, with a region in the center of two to four millimeters where binocular vision obtained.[18]

[17] The Ophthalmograph is manufactured by the American Optical Co.

[18] This test (bar reading) is used in orthoptic training to strengthen binocular fusion.

The paragraphs ranged from 13 to 17 lines in length. Each line was 10.2 cm long (approximately 17 degrees at the eye) and printed in 12 point type. The illumination on the reading material was 40 foot-candles.

No attempt was made to equate the paragraphs for difficulty, since this depends to a large extent on the background of the subject rather than on the paragraph itself. The paragraphs were rotated in the procedure so as to minimize differences in difficulty and practice effect. As may be seen in the following scheme, each paragraph was read only once by each subject—a different paragraph for each of the four conditions. The paragraphs were rotated so that each paragraph was read equally often under the four conditions. In addition each subject in successive groups of four started reading with a different condition indicated by the italicized number, so that the initial paragraph was read equally often under the four conditions proceeding clockwise.

Subject	Binocular	Alternating Monocular	Monocular	
			Right	Left
1.	*1*	2	3	4
2.	1	*2*	3	4
3.	1	2	*3*	4
4.	1	2	3	*4*
5.	*2*	3	4	1
6.	2	*3*	4	1
7.	2	3	*4*	1
8.	2	3	4	*1*
9.	*3*	4	1	2
10.	3	*4*	1	2
11.	3	4	*1*	2
12.	3	4	1	*2*
13.	*4*	1	2	3
14.	4	*1*	2	3
15.	4	1	*2*	3
16.	4	1	2	*3*

The usual measurements of time in seconds, number of fixations and regressions per line were made from the photographic film. Characteristic patterns are illustrated. Comprehension on each of the paragraphs was tested at the end of reading four paragraphs by answering five written true-false questions on each paragraph. The purpose of instructing the subject that he would be required to answer questions on the paragraphs read, was to insure, as far as possible, an understanding reading of the material.

Short Exposure Span for Meaningful Material.—After the eye movement records had been obtained, the subject was seated before a focal plane tachistoscope (speed 1/90 of a second) with illumination and reading distance the same as for the eye movement records. The material for this test was a practice set of five cards and three sets of ten cards upon which had been mounted paragraphs similar to those used for the eye movement photographs. The object here was to make the test for span similar to the conditions encountered in reading. After the practice set had been presented so that the subject understood that he was to report all the words, parts of words, letters, punctuation, etc., seen on one or more lines, a set was presented for both eyes and for each eye separately. The material and procedure were rotated as for the eye movements.

The span was measured in letter spaces. It was the total horizontal distance along a line of print within which all the words, letters and punctuation marks were reported correctly. If words were reported from two adjacent lines, the span was the horizontal distance within which all the words, etc., in one line or the other were reported correctly. For example, if the subject correctly reported

The best . . .
. . . recently, ...

the span was taken to be 14 letter spaces.

Paracentral Visual Acuity.—By means of the short exposure technique, the increase of errors in seeing a standard test object when moved peripherally on either side of the fixation point was measured for the two eyes and for each eye separately. For the present purpose the measure might be called an index of relative acuity.[19] A Whipple Tachistoscope was used for this work, set to operate at 1/75 of a second with an illumination on the material of 20 foot-candles.

A series of 68 test cards was presented. On each of these appeared one Landolt Ring. The rings were so placed on the cards that they appeared in seventeen different positions on the horizontal, right and left of and on the fixation point. These positions were 5 millimeters apart, up to 4 centimeters right and left of the fixation point. The ring appeared four times in each position with the opening either right, left, up or down, thus making a total of 68. The order of presentation was determined by chance. The distance of the eyes from the material exposed was 80 centimeters while the fixation point was two centimeters in front of the material. The extreme left and right positions were 2 degrees and 52 minutes of visual angle from the fixation point. The Landolt rings subtended an angle of 17 minutes of arc and were therefore several times larger than the size of the test letter which can be read with normal foveal acuity.

The instructions to the subject were to regard the fixation point and after each exposure to report whether the opening in the ring was right, left, up or down. The scoring was the number of errors in each position to the right and left of the fixation point.

Preliminary Experiment—Results

In a preliminary experiment, eye movements while reading were obtained with twelve subjects using the procedure outlined above with the exception that rotation of paragraphs and conditions was not used. The results for two reading measures are shown below. The changes would seem to support the contention that reading with monocular vision for short periods is significantly favorable to reading speed. On the other hand, the changes could just as well have resulted from a practice effect or from differences in difficulty between

	Binocular	Monocular		
		Alternating Monocular	Right	Left
Time per line (seconds) Mean.....	1.61	1.51	1.43	1.38
Fixations per line Mean.....	6.84	6.54	6.32	6.10

the paragraphs. On an a priori basis, the increase in speed and decrease in fixation frequency appeared to be larger than one might expect from practice effect or from the paragraphs themselves.

These preliminary results therefore gave impetus to the experiment precisely as outlined in the procedure, the results of which follow. The experiment was started anew with different subjects since subjects new to the conditions of the experiment were required.

[19] This is to be distinguished from visual acuity as the term is used in ophthalmology.

Main Experiment—Results

Eye Movement Measures.—The mean reading time per line under the four conditions of the experiment was substantially the same for the group as a whole (Table 2 and Fig. 1). There was no difference in the speed of reading as between binocular reading and reading monocularly with either eye, or alternately. In addition the variability remained relatively constant.

TABLE 2

Mean Time per Line (Seconds) for Binocular and Monocular Reading

Subjects	Binocular	Monocular		
		Right and Left Alternately	Right	Left
Mos	1.53	1.86	1.49	1.93
Kar	2.34	3.25	2.51	2.06
Alt	1.65	2.02	1.98	2.02
Rak	1.97	2.02	2.36	1.93
Fle	2.19	1.96	1.76	1.67
Sew	1.92	1.85	1.89	1.82
Wil	3.37	3.79	3.56	3.45
Goo	1.49	1.27	1.18	1.26
Pet	1.77	1.98	1.46	2.03
Dud	1.92	1.93	2.27	1.85
Tay	2.11	2.10	2.36	2.65
Kun	1.65	1.59	1.91	1.96
Het	1.05	1.05	.95	.97
Mor	1.59	1.67	1.92	1.68
Rha	1.37	1.41	1.63	1.51
McF	1.73	1.59	1.52	1.45
Mean	1.85	1.96	1.92	1.89
$SD_{dist.}$	.51	.66	.60	.55

A comparison of the mean number of fixations recorded for binocular reading, and for the right and left eyes alternately and with the right eye alone showed differences which were slight and not significant (Table 3 and Fig. 1). A somewhat larger decrease in the number of fixations for reading with the left eye alone may be noted, but the difference is not significant. The variability is not constant over the four conditions but indicates no particular trend.

On the average, the changes in the number of regressions (Table 4) in the four conditions are small and not significant. There is a slight decrease in variability for the right and left eyes.

If we compute the approximate pause duration (average time taken by the fixations) we find an increase of slightly more than five percent (not significant) on the average for the monocular reading when compared to the binocular:

	Seconds	Δ%
Binocular	.227	—
Right and left alternately	.240	5.8
Right	.238	4.9
Left	.241	6.1

There is a tendency (not significant) for an increase in comprehension (Table 5) when reading with separate eyes. The increase is not of constant amount for the three monocular conditions. The indications are that the paragraphs were read understandingly since the averages are close to four out of five correct answers to the questions. The main purpose of this test was to insure that reading did not degenerate into skimming with no attention to the material.

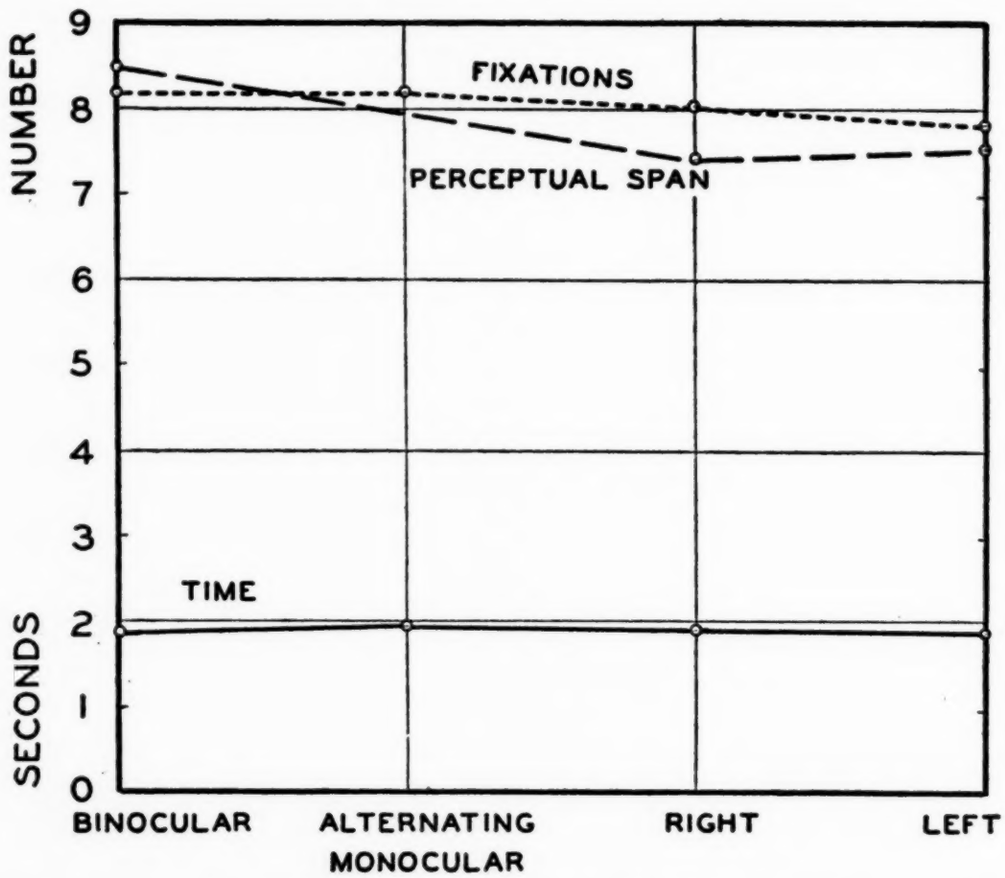

FIG. 1. Comparison of reading time, fixation frequency per line, and perceptual span for monocular and binocular vision.

The questions were sufficiently specific to insure that these answers were largely the result of comprehending reading. Apparently, no disadvantage in comprehension has been introduced by forced monocular reading for short lengths of material.

As a whole, the eye movement measures are in agreement in revealing no apparent increase or decrease in the efficiency of reading for short lengths of material when monocular vision is used. Thus no valid support is shown for the hypothesis that reading is more efficient with monocular vision under the conditions of this experiment.

TABLE 3

MEAN NUMBER OF FIXATIONS MADE PER LINE DURING BINOCULAR AND MONOCULAR READING

Subjects	Binocular	Monocular		
		Right and Left Alternately	Right	Left
Mos.	7.4	9.4	7.5	9.8
Kar.	9.9	12.2	10.0	7.4
Alt.	7.6	9.1	7.8	7.8
Rak.	8.3	7.6	8.1	6.8
Fle.	10.3	9.1	8.1	8.7
Sew.	7.4	7.2	7.3	7.3
Wil.	13.3	11.7	11.9	10.5
Goo.	6.6	6.2	5.8	6.3
Pet.	6.9	8.0	6.4	8.3
Dud.	9.1	8.9	10.9	9.9
Tay.	8.9	8.5	9.1	8.6
Kun.	7.6	6.5	8.2	8.1
Het.	4.8	4.9	4.7	4.7
Mor.	8.0	7.8	9.1	7.7
Rha.	5.9	6.3	7.1	6.6
McF.	8.2	7.5	7.0	6.8
Mean.	8.14	8.18	8.06	7.83
$SD_{dist.}$	1.89	2.05	2.85	1.45

TABLE 4

MEAN NUMBER OF REGRESSIONS MADE PER LINE DURING BINOCULAR AND MONOCULAR READING

Subjects	Binocular	Monocular		
		Right and Left Alternately	Right	Left
Mos.	1.4	3.2	2.2	2.7
Kar.	2.8	3.6	2.1	2.2
Alt.	1.2	2.2	.3	.6
Rak.	1.1	1.1	1.3	.7
Fle.	2.9	2.1	1.8	2.7
Sew.	1.7	1.3	1.8	2.0
Wil.	4.7	3.6	4.0	3.2
Goo.	1.2	.8	.6	.9
Pet.	.7	1.6	.8	1.6
Dud.	2.0	1.9	2.8	1.6
Tay.	1.6	1.3	1.7	.8
Kun.	.9	.7	1.5	1.5
Het.	0.0	.1	.3	.1
Mor.	1.2	.9	1.7	.8
Rha.	.7	.7	.8	.7
McF.	1.4	1.0	.5	.6
Mean.	1.59	1.64	1.51	1.42
$SD_{dist.}$	1.08	1.03	.96	.88

Perceptual Span for Meaningful Material.—An average decrease in the perceptual spans of 12 and 10.5 percent was found for the right and left eyes respectively as compared to the span for binocular vision (Table 6). These decreases may be considered statistically significant, exceeding Fisher's[20] 5 percent point of significance for the right and falling short of this point by .04 for the left eye.[21] In addition the variability for the monocular perceptual spans showed a slight increase. The correlations between low fixation frequency and large

TABLE 5

COMPREHENSION ON READING PARAGRAPHS

Subjects	Binocular	Monocular		
		Right and Left Alternately	Right	Left
Mos	5	4	3	4
Kar	4	5	4	5
Alt	3	5	5	4
Rak	4	5	2	2
Fle	4	4	4	3
Sew	5	5	3	5
Wil	5	5	5	4
Goo	5	2	5	3
Pet	4	4	5	5
Dud	4	5	2	5
Tay	2	5	4	5
Kun	2	2	4	5
Het	4	2	5	4
Mor	4	3	5	4
Rha	2	3	3	5
McF	3	3	4	5
Mean	3.75	3.88	3.94	4.25
$SD_{dist.}$	1.03	1.15	1.02	1.44

perceptual span, which should be intimately related (according to our second hypothesis), are as follows: for binocular vision $r = +.19$, for the right eye $r = +.49$, and for the left eye $r = +.26$. The slight or moderate relationship shown here does not lend any real support to our hypothesis, for if the individual were reading to the limits of his span, the correlation between the span and fixation frequency should be high.

In addition the above decrease in span should be reflected in an increase in the number of fixations, if the hypothesis were correct. The relations between span, time, and fixation frequency are shown

[20] Fisher, R. A. *Statistical methods for research workers.* Edinburgh: Oliver & Boyd, 1933.

[21] Significances of differences were determined from actual differences between the series. The critical ratio is the ratio of the mean of the differences to its standard error, using $\sqrt{N-1}$ in computing the standard error. See Guilford, J. P. *Psychometric methods.* New York: McGraw-Hill, 1936.

in Fig. 1. There we find that the decrease in span in monocular vision has not resulted in a concomitant increase in reading time or fixation frequency. As a matter of fact the fixation frequency showed a very slight tendency to decrease.

Thus the data on spans as related to reading efficiency are evidence against the view that perceptual span is intimately related to reading efficiency, and especially against the view that the span limits reading performance.

TABLE 6

SPAN IN LETTER SPACES FOR MEANINGFUL MATERIAL (FOCAL PLANE TACHISTOSCOPE)

Subjects	Binocular	Monocular	
		Right	Left
Mos.	8.2	7.6	9.4
Kar.	9.8	5.2	6.2
Alt.	11.4	10.2	9.0
Rak.	9.1	6.4	5.5
Fle.	7.8	7.0	7.1
Sew.	7.3	8.1	5.2
Wil.	8.6	5.6	7.9
Goo.	10.9	11.4	9.7
Pet.	11.3	11.2	10.5
Dud.	7.0	5.3	6.0
Tay.	6.3	6.3	5.3
Kun.	5.7	4.1	5.5
Het.	9.7	9.4	10.1
Mor.	5.6	4.6	5.0
Rha.	7.7	8.7	10.7
Mean.	8.43	7.41	7.54
$SD_{dist.}$	1.83	2.25	2.09
CR.	—	2.68	2.10

Paracentral Visual Acuity.—In connection with the superiority of the binocular perceptual span, it is important to know whether the area of adequately clear vision shows a similar monocular-binocular difference. The falling away of accurate vision (as shown by errors in the Landolt Ring test), when the test object is moved slightly to the right or left of the fixation point, is shown in Fig. 2 and Table 7. Binocular vision is definitely superior in this test. The percent of errors is smaller in binocular vision for every distance from the fixation point. The errors reach the 25 percent level at different distances in the different curves. If we take the distance along the 25 percent level, between the two crossings of the same curve, as a measure of the area available for a certain grade of clear vision, we find this area smaller for the separate eyes than for both eyes. This same difference can be found in the totals (Table 7) where the errors for the separate eyes increase by 7.7 and 10.0 percent of the possible

errors. (By 54 and 70 percent of the binocular errors, a reliable difference with critical ratios of 3.05 and 4.12.)

The correlation between errors on the test of paracentral acuity and fixation frequency while reading under similar conditions are as follows: for binocular reading rho = +.22, for the right eye rho = +.48 and for the left rho = +.10. There is thus a moderate tendency for a low fixation frequency to be associated with a relatively high acuity in paracentral vision.

It is apparent that we have the same tendency here for a restriction of the area of clear vision on the horizontal when we compare the

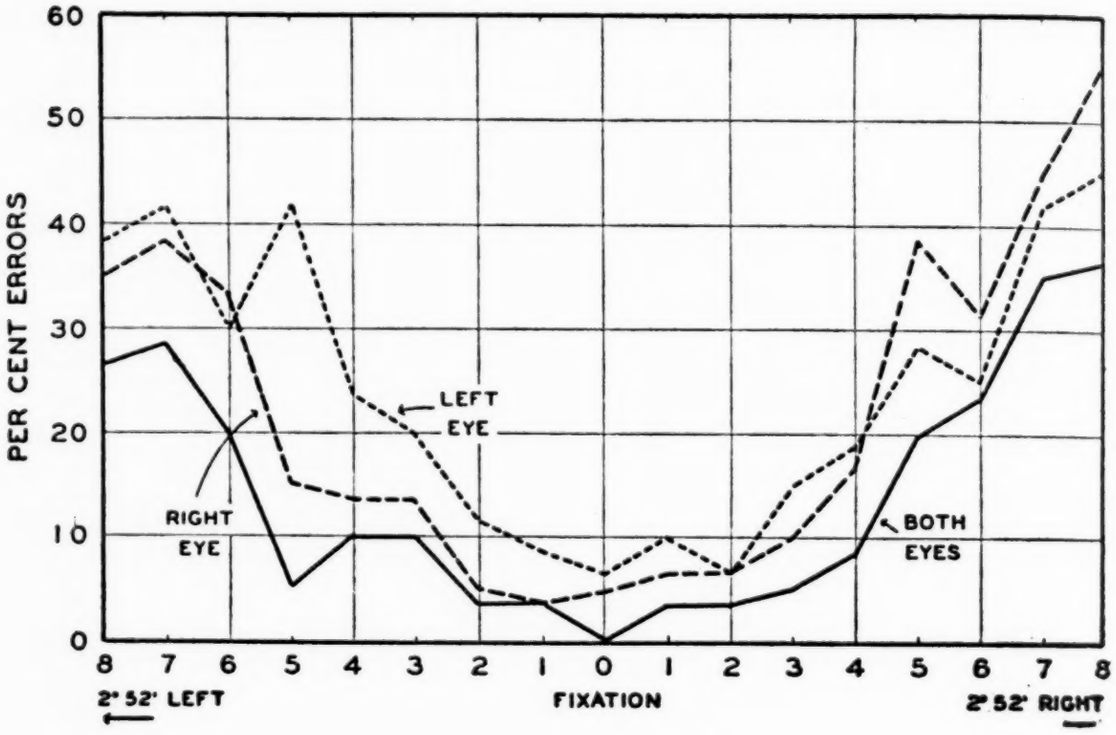

FIG. 2. This chart shows the increase in errors for both and separate eyes as the test ring is moved away to the right and left of the fixation point.

monocular to the binocular as found above in the span for meaningful material. It will be recalled that these differences are significant. Yet the calculated correlation coefficient on the relation between span for meaningful material and errors on the test of paracentral visual acuity are +.07, +.55 and +.07 for the three conditions. In spite of this, the narrowing of the perceptual span with monocular vision would seem to find its most promising explanation in terms of acuity of vision. It is obvious, however, that the data bearing on this question are not conclusive.

It will be noted that the curve for the left eye lies farther to the right than the curve for the right eye. This shift is actually quite small when the average for the curve is taken and is of the magnitude

which might be expected since the fixation point was two centimeters in front of the material to be exposed. If the line of regard of the left eye be carried back of the fixation point to the material, it would interesect the plane of the card to the right of the center (point on the card directly behind the fixation point).

TABLE 7

SUMMARY OF ERRORS MADE IN EACH POSITION OF PRESENTATION (TACHISTOSCOPE) FOR NON-MEANINGFUL MATERIAL. FOUR CORRECT RESPONSES POSSIBLE IN EACH POSITION

Both Eyes

		Position																	Total Errors
		Left						Fixation					Right						
		8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	
1	Mos	2	1	2	1	2	1	1	0	0	1	1	0	1	2	3	2	2	22
2	Kar	2	3	0	1	1	1	0	1	0	0	0	1	1	0	2	2	0	15
3	Alt	0	1	2	0	0	0	0	0	0	0	0	0	0	1	0	1	1	6
4	Rak	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	1	4
5	Fle	1	1	0	0	0	0	0	0	0	0	0	0	0	2	2	1	1	8
6	Sew	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	4
7	Wil	4	2	3	1	0	0	0	0	0	0	0	0	0	0	0	2	2	14
8	Goo	1	1	0	0	1	3	0	1	0	1	0	0	1	1	1	2	2	15
9	Pet	2	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	5
10	Dud	1	1	1	0	1	1	0	0	0	0	0	0	0	1	0	1	3	10
11	Tay	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	6
12	Kun	1	2	2	0	1	0	0	0	0	0	1	1	1	2	2	2	3	18
13	Het	0	0	1	0	0	0	0	0	0	0	0	0	0	0	2	2	1	6
14	Mor	0	1	1	0	0	0	0	0	0	0	0	1	1	0	1	0	1	6
15	Rha	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	2	2	6
Total		16	17	12	3	6	6	2	2	0	2	2	3	5	12	14	21	22	145
% Errors		26.7	28.3	20.0	5.0	10.0	10.0	3.3	3.3	0	3.3	3.3	5.0	8.3	20.0	23.3	35.0	36.7	14.2

TABLE 7 (continued)

Right Eye

		Position																	Total Errors
		Left						Fixation					Right						
		8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	
1	Mos	2	3	2	1	1	1	1	0	0	2	0	0	1	1	3	1	2	21
2	Kar	4	2	2	2	2	2	3	1	1	1	2	3	3	3	3	4	3	40
3	Alt	1	1	1	1	0	0	0	0	0	0	0	1	1	1	0	0	2	9
4	Rak	2	2	1	2	0	0	0	0	0	0	1	0	0	1	0	3	1	13
5	Fle	0	1	2	0	1	0	0	0	0	0	0	1	0	2	2	3	1	13
6	Sew	1	1	2	0	0	1	0	0	0	0	0	0	0	3	2	2	2	14
7	Wil	3	1	1	0	2	1	0	1	0	0	0	0	0	0	1	2	2	14
8	Goo	0	2	1	0	1	0	1	0	0	0	0	0	0	1	1	4	3	14
9	Pet	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2	5
10	Dud	1	0	1	1	0	1	0	0	1	0	0	0	1	2	2	3	4	17
11	Tay	2	2	1	1	0	0	0	0	0	0	0	0	0	1	0	1	3	11
12	Kun	2	3	3	1	0	1	0	0	1	1	1	1	2	3	2	2	3	26
13	Het	2	1	1	0	1	0	0	0	0	0	0	0	0	1	1	1	1	9
14	Mor	1	1	1	0	0	0	0	0	0	0	0	0	2	2	1	0	3	11
15	Rha	0	1	0	0	0	0	0	0	0	0	0	0	0	2	1	1	1	6
Total		21	23	20	9	8	8	3	2	3	4	4	6	10	23	19	27	33	223
% Errors		35.0	38.3	33.3	15.0	13.3	13.3	5.0	3.3	5.0	6.7	6.7	10.0	16.7	38.3	31.7	45.0	55.0	21.9

TABLE 7 (continued)

Left Eye

		Position																	Total Errors
		Left						Fixation					Right						
		8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	
1	Mos	2	3	1	2	0	1	0	0	0	0	0	0	0	2	2	2	3	18
2	Kar	2	1	2	3	3	3	1	1	2	2	0	1	1	2	1	2	2	29
3	Alt	1	2	2	1	2	0	0	0	0	1	0	0	0	2	1	2	3	17
4	Rak	1	1	2	2	0	0	0	0	0	0	0	0	0	0	0	2	1	9
5	Fle	1	2	1	1	0	1	0	0	0	0	0	1	0	0	1	3	0	11
6	Sew	0	1	1	0	0	0	0	0	0	0	0	0	1	1	1	0	1	6
7	Wil	3	3	1	2	1	0	0	0	0	0	0	0	0	1	0	1	0	12
8	Goo	3	2	2	4	2	2	3	2	1	2	0	2	2	0	2	0	3	32
9	Pet	1	2	1	1	1	0	0	0	0	0	0	1	0	1	1	1	2	12
10	Dud	2	0	0	1	1	1	0	1	0	0	0	0	1	3	2	3	2	17
11	Tay	0	2	0	3	0	1	1	1	0	0	0	1	0	0	0	1	2	12
12	Kun	3	3	1	2	2	1	0	0	0	0	2	1	3	2	2	2	2	26
13	Het	1	2	2	1	0	0	0	0	0	0	0	0	0	0	0	1	2	9
14	Mor	3	1	2	2	1	2	2	0	1	1	2	1	3	0	2	1	1	25
15	Rha	0	0	0	0	1	0	0	0	0	0	0	1	0	3	0	4	3	12
Total		23	25	18	25	14	12	7	5	4	6	4	9	11	17	15	25	27	247
% Errors		38.3	41.6	30.0	41.6	23.3	20.0	11.7	8.3	6.7	10.0	6.7	15.0	18.3	28.3	25.0	41.6	45.0	24.2

Eye Movements—Descriptive.—In Fig. 3 are presented several sections of eye movement photographs obtained under the four conditions of the experiment. As usual, the binocular record shows two traces, one for each eye. A similar record is obtained for reading with the right and left eye alternately, since the bar was in such a position that it did not obstruct the reflexes from the two eyes. The necessity of covering an eye for reading with the right or left singly, resulted in a record with a single trace. It is important to point out that these illustrations are generally typical of the series of records as a whole, with regard to the points under discussion in this section.

The most important evidence from a study of the general pattern of the records for moncular and binocular reading is the presence of vergence movements in monocular reading when binocular fusion is absent. Reading the records downward, one sees, for the initial fixation on the complete line of the binocular record, the typical narrowing of the distance between the traces during the fixation, after a previous widening from the end of the preceding fixation. In the record for alternating monocular vision, the first fixation on the complete line shows similar activity of the eyes, yet the left eye alone was reading the first half of the line. It is apparent from the single traces in the single eye records that similar behavior of the eyes is again taking place. These vergence movements are present to a lesser degree in succeeding fixations of the lines, as is usually the case with binocular reading. Since a direct stimulus to binocular fusion is absent in monocular reading, these vergence changes evi-

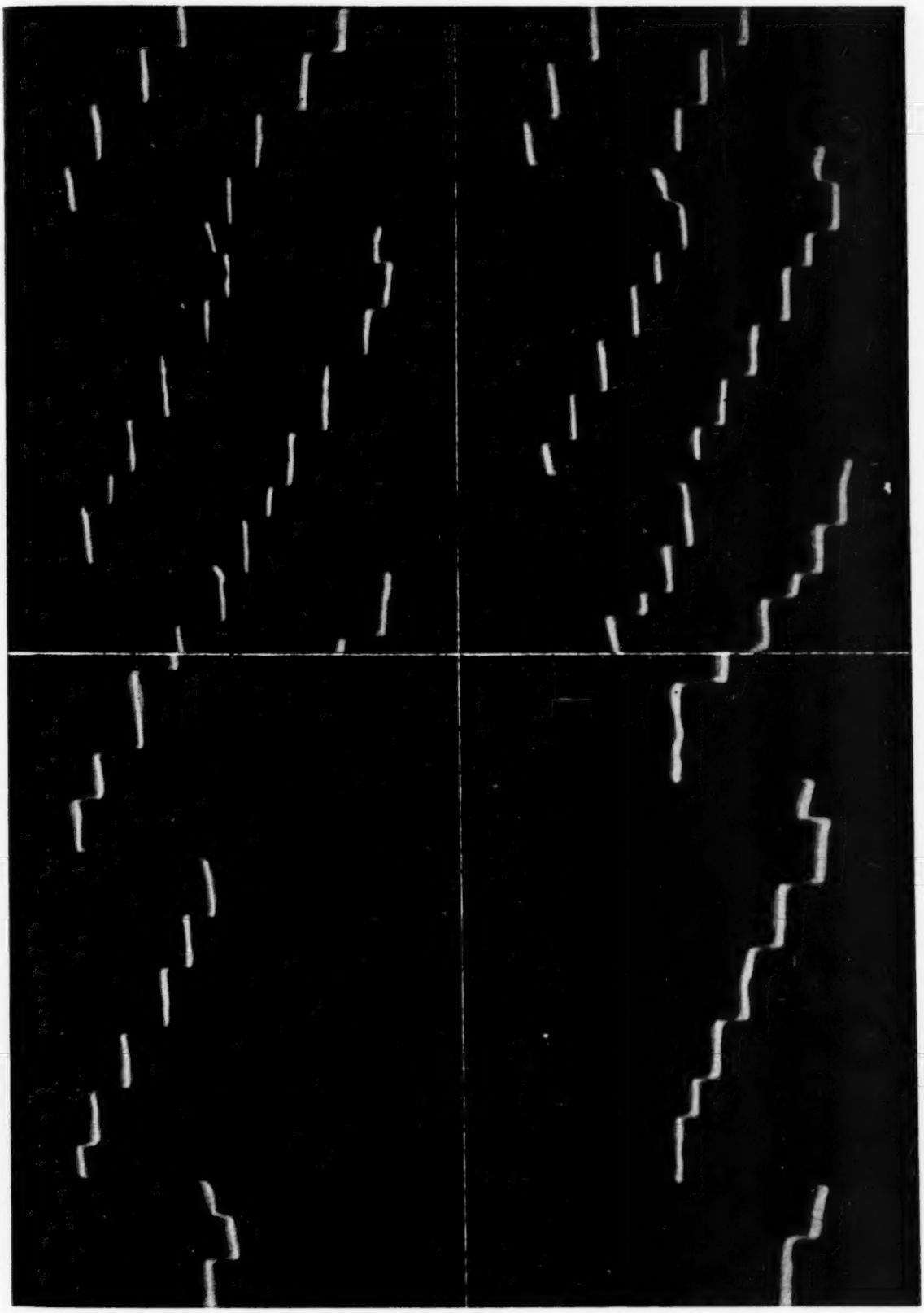

FIG. 3. Top left, binocular reading; top right, alternating monocular; lower left, left eye alone; and lower right, right eye alone (the records should be read downwards).

dently depend upon changes in accommodation, as a result of the intimate relation of convergence and accommodation.

Discussion

At the beginning of the present investigation two hypotheses were proposed for testing: (1) that by dispensing with the necessity for binocular fusion [*i.e.*, by using monocular vision] reading efficiency should be materially increased; and (2) that perceptual span is a limiting factor in reading efficiency and should therefore be intimately related to fixation frequency in reading.

The experimental results yielded no material support for either contention. Reading efficiency was neither increased nor decreased by monocular as compared with binocular reading, as shown by the relatively constant average values of reading time and fixation frequency for those conditions. Apparently, the forced use of a single eye, for short lengths of material, introduces no handicap in the reading process. But it does not produce any increase of reading efficiency. Dispensing with the necessity of binocular fusion with relatively normal adult eyes, offers no advantage in reading. No motive from this source exists for resorting to monocular reading when the individual is being trained in rapid reading. Since the reader derives no advantage from monocular reading, there would seem to be nothing in the situation to force him to the use of a single eye.

The discovery that monocular perceptual spans for meaningful material were significantly smaller on the average than the binocular spans, makes it surprising that, instead of the expected increase in reading efficiency for monocular vision, there was no proportionate decrease in the level of the reading performance expecially as measured by the fixation frequency. On an a priori basis, the smaller span, if a limiting factor, would require a greater fixation frequency. In addition the correlations between large perceptual span and low fixation frequency though positive, were much too low for purposes of prediction. By knowing the relative sizes of the perceptual span for a group of individuals, one could not predict their relative standing in reading efficiency. Therefore, it appears that perceptual span is not a limiting factor in the usual reading performance. Obviously, there must be some relation of perceptual span to the reading process, but the evidence makes it apparent that full use of the span is not made in reading under the conditions of this experiment.

It is difficult to imagine that the monocular decrease in perceptual span for meaningful material is due to any other factor than an actual decrease in material seen—that is, to a retinal limitation.

The data on paracentral acuity offer evidence to support this suggestion. There was a ten percent increase (significant) in errors of seeing by the single eye. Apparently, the area available for any particular grade of seeing is smaller on the average for one eye than for both, and would result in less material seen in the monocular test of perceptual span. (These measurements were within a radius of 3° of the center of the fovea.)

Taken as a whole, the data on perceptual spans indicate that the span is probably limited by acuity, but that span does not in turn limit reading efficiency. If this is so, then attempts to increase perceptual span for the purpose of increasing reading efficiency would seem to invite failure from two directions. First, acuity probably cannot be materially increased. Second, increasing the span (so far as possible) would not guarantee an increase in reading efficiency, since span is probably not a limiting factor.

In this study the vergence movements in the eye movement records for monocular vision were similar in pattern and extent to those found in binocular reading, occurring when the direct stimulus for fusion of the eyes was absent. The close association of accommodation and convergence makes it more than likely that accommodation plays the dominant role in the ocular activity manifested in vergence movements during monocular reading. It is extremely easy to place the responsibility upon accommodation for the same behavior while reading with both eyes. The suggestion that accommodation plays a large part in coordinate movements of the visual axes in reading has been offered previously by McFarland, Knehr and Berens [22] to explain the decrease in magnitude of these movements in oxygen lack, and the appearance of the same movements in patients with heterotropias (squint). The evidence from the present investigation makes this explanation very probable.

If this interpretation of the vergence movements occurring during reading approximates the true state of affairs, we can describe the vergence movements in terms of accommodation somewhat as follows. An increase of accommodative effort produces simultaneously an increase in the convergence of the visual axes. At the beginning of a fixation, accommodative effort is greatest, convergence greatest, and eye traces on the photographic record farthest apart. During the fixation accommodation relaxes, convergence relaxes, and the traces on the eye movement record move closer together. This series of events is repeated for each fixation, although it is usually most marked in the initial fixation on a line.

The movement of the visual axes relative to each other during fixation is relatively large, amounting to from a half to more than a

[22] *Op. cit.*

degree of arc. Measurements on the longitudinal magnitude of corresponding fusional areas give us additional evidence on the interpretation of these large vergence movements. Surrounding the pairs of "corresponding retinal points there are effective areas of sensation within which similar images will be fused and will appear binocularly as one." [23] The angular magnitude of these areas increases as they are more peripheral, varying from roughly 3 minutes of arc at the fovea to 9 minutes of arc at a peripheral angle of 8 degrees for one individual. Outside of these boundaries of corresponding fusional areas points will be seen double, although frequently one or both may be suppressed. It becomes immediately apparent that the relative angular movements of the visual axes during fixation while reading are many times the angular size of corresponding fusional areas. It follows that single binocular vision cannot obtain for more than a fraction of the total time of fixation. At some time during the fixation word images are seen singly and accommodation is optimal. All the rest of the time, when the images should be seen double, they are either suppressed or ignored, as visual impressions are ignored during the saccadic movement between fixations. There seems to be no particular reason, anyway, for taking one-fifth of a second to sense an amount of reading material which may be easily seen in an exposure of one-fiftieth of a second. Apparently, the visual mechanism is not taxed in reading. Rather, it seems to idle while the perceptual processes are in operation, being at all times, however, at their beck and call. In all probability, the visual apparatus is capable of feeding material to the centers faster than it can be handled by the perceptual processes.

It seems evident from the data of this study, in the light of work by other investigators, that measurement of vergence movements in an eye movement record, or analysis of different patterns of normal oculomotor behavior, will yield little information of any consequence in the interpretation of reading efficiency. In a purely practical sense, it would seem to be cheaper and easier for the schools to use written reading tests instead of eye movement records for their usual work. As we have noted above, several photographic measures of reading efficiency are highly valid and reliable, but so are certain of the reading tests. The technique of eye movement photography will continue to be an important tool for experimental investigation of processes which are related in any way to oculomotor behavior.

Although the results of this study may not be considered conclusive on any point, they do provide additional evidence for the

[23] Ogle, K. H. Analytical treatment of the longitudinal horopter; its measurement and application to related phenomena, especially to the relative size and shape of the ocular images. *J. opt. Soc. Amer.*, 1932, **22**, 665.

views of such investigators as Tinker and Stromberg who maintain that the important limiting factors in reading are central rather than peripheral.

Further data on this general problem might be obtained from longer periods of monocular reading, with photography of the oculomotor behavior. Children of primary school age, many of whom are hyperopic, might be fit subjects for a repetition of the general features of the present experiment. The role of accommodation in the coördinate activity of the two eyes might be further studied by instilling drugs for relaxation of accommodation, and noting the changes, if any, in the pattern of eye movements during reading. It would be necessary for lenses to be worn while reading under these conditions in order that clear seeing might obtain. In addition, fatigue is an unknown factor running through these problems.

Summary and Conclusions

This study compared, on a group of 16 normal adult subjects, the effects of reading with monocular and binocular vision on eye movement measures of reading efficiency for short lengths of material. Correlative data were obtained in the form of perceptual spans for monocular and binocular vision on material similar to that used in reading. A span for non-meaningful material was also determined by means of the decreasing accuracy of seeing a standard test object moved right and left of the fixation to a distance of approximately 3 degrees. The main conclusions are as follows.

1. No evidence was obtained to support the hypothesis that reading efficiency should increase when binocular fusion is dispensed with [*i.e.*, in monocular vision]. Monocular reading time, fixation frequency, and regression frequency did not differ significantly from the binocular.

2. Perceptual spans for monocular vision showed a significant decrease of ten percent as compared to binocular perceptual spans. A similar decrease was found in the paracentral area of clear vision for non-meaningful material, in monocular compared to binocular vision.

3. No increase in the number of fixations (or alteration in reading efficiency) proportional to the decrease in perceptual span was noted for monocular reading. Low coefficients of correlation were obtained between large spans and small fixation frequency. The evidence indicates that perceptual spans are not a limiting factor in the usual reading performance.

4. The vergence movements noted in binocular eye movement records retain their characteristic form and approximate magnitude

in monocular reading. This may be considered to be additional evidence that changes in accommodation account largely for these changes in vergence.

5. Considered as a whole, the data provide further evidence that reading efficiency depends largely upon the individual's familiarity with words as symbols and his store of ideas and meanings. Efficiency of reading seems to depend to a large extent upon central rather than peripheral factors.

(Manuscript received January 27, 1941)

MIRROR POSITION AND NEGATIVE TRANSFER

BY THOMAS W. COOK

Acadia University

An experiment by Hall (**3**) has shown that after subjects had learned to trace a maze seen indirectly in a mirror placed in one of three positions (to the right, centrally, or on the left side of the subject), shifting the mirror to another of these positions resulted in negative transfer. As Hall suggests, her results leave unsolved the problem of the nature of the inhibitory effect. The concept of fatigue (applying the term in its most objective sense to signify a drop in efficiency with constant stimulus situation) does not aid in integrating the data with results of previous experiments. For we would have to assume that fatigue was present during the first 10 trials but was masked by learning; and furthermore that the fatigue transferred to a second mirror position while the learning did not. The later conclusion, at least, is not in harmony with the data on transfer of learning and fatigue.

However, the question of short time ('temporary') versus long-time 'semi-permanent') inhibitory effects seemed worth investigating. Accordingly, two of the writer's assistants (Miss Ellen Piers and Mr. Stanley Borden) duplicated Hall's experiment with a single change in procedure: each trial in a series of 10 trials with a given mirror position was separated by a 30 second interval, and a 5 minute rest period was inserted between the end of each series of 10 trials and a second or third series with a new mirror position.

The results for time (errors were not measurable with the technique used) show marked positive rather than negative transfer for the first two trials, from mirror position 1 to mirror position 2. Trials 1 and 2 take 39.6 and 26.7 seconds with position 1, and 27.8 and 22.1 seconds with position 2. For the remaining trials there is either no difference or a slight negative transfer from position 1 to position 2, and a consistent though slight (about 5 percent) negative transfer from position 2 to position 3.

So far the experimental data could be taken to indicate that the inhibitory effect could be minimized, if not entirely eliminated, by spacing trials farther apart, and the investigation might have been suspended at this point had it not been for an acute observation by Mr. Borden (later supported by evidence from watching and questioning subjects, and by having the latter attempt to trace the maze blindfolded) to the effect that many subjects had developed a method

of tracing that to some extent 'avoided' the reversing effect of mirror position. This method, as far as I can generalize from its description by individual subjects, consisted in paying attention to kinæsthesis during initial learning, thus building up a motor pattern that enabled them to trace the maze with little dependence upon vision. We thus have to reckon with the possibility that the inclusion of subjects tracing by the motor method might be masking the inhibitory effect of earlier upon later mirror positions with visually minded subjects. Also, the inclusion of a larger number of motor-minded subjects in the Piers-Borden (as compared with Hall's groups) might account for the partial shift to positive transfer in their data.

Since it was not feasible to separate subjects into motor and visual categories by an a priori criterion, it was decided to attempt to construct a pattern that could not be traced successfully by the motor method. Double lines one-eighth inch apart, bounding a path roughly 14 inches long, were drawn on a stencil and printed on a multigraph machine. The chief aim in constructing the pattern was to make it as irregular as possible. Since the particular pattern is of no importance it is not reproduced here, but copies will be furnished on request.

The mirror tracing technique was similar to that used with the familiar star-shaped pattern of laboratory courses in psychology. The subject was requested to trace the pattern from beginning to end, paying attention both to accuracy and speed. He was told that an error consisted in touching the side of the path, that he must not lift the pencil while tracing, and that if he went outside the path he was to reënter it as near the point of departure as possible. The self-feeding pencil used in tracing was provided with a finger grip by soldering a wire collar to the barrel two inches from the pencil point. Exploratory work showed that 10 trials with each mirror position took more time than some subjects would readily give. Accordingly only 5 trials were traced with each position of the mirror. The time between trials was 2 minutes; a 5 minute interval separated each group of 5 trials. The mirror tracing apparatus was a board screened from the subject's direct vision. On this board a copy of the maze was fastened with thumb tacks. In the interval between trials the used copy was removed and an unused one substituted by the experimenter. Time intervals to the nearest second were recorded with stop clocks. The subjects were students at Acadia University.

Because of the unambiguous nature of the results only 3 groups of subjects were tested. The order of mirror positions was CLR, LCR, RCL for Groups I, II, and III respectively. Groups I and II thus admit a comparison of the first and second 5 trials (combined results of CL and LC), and Groups I and III a comparison of the first 5 trials and the third 5 trials (combined results of LR RL).

Results

The error data appear in Table I, and indicate marked negative transfer. All trials show progressively more errors from position 1 to position 2 and position 3. Taking the average errors at position 1 to equal 100, the average errors for position 2 is from 140 percent to 143 percent (by different methods of computation), and for position 3, from 173 percent to 177 percent. For time the results are not so clearly indicative of interference. The average speed of tracing in position 1 and position 3 are nearly identical, while position 2

TABLE I

Average Errors at Different Mirror Positions

Trial	Groups I and II		Groups I and III	
	M.P. 1 (CL)	M.P. 2 (LC)	M.P. 1 (RL)	M.P. 3 (LR)
1	38.5	62.5	28.3	51.5
2	19.9	36.2	23.8	33.2
3	18.4	23.4	14.0	25.9
4	15.8	18.2	11.0	20.8
5	15.9	18.5	11.3	21.8

requires an average of 168 seconds per trial as compared with 148 seconds for position 1. The time data therefore show (as was suggested also by qualitative results) that there are positive as well as negative features to the transfer situation, and while for the error measure the negative factors strongly outweigh the positive, for time the negative factors cause a drop in relative efficiency only with position 2, after which they begin to be neutralized by positive factors. It should of course be noted that the time measures in the Piers-Borden and the present experiment are not comparable, since in the latter subjects were told to try to avoid errors.

Cross Education

To check the possibility that the contrast between negative transfer with shift in mirror position as against positive transfer in cross education (**2**) might be to some extent a function of the maze pattern learned in the present experiment, 10 subjects were given 5 trials with the right hand followed by 5 trials with the left hand, and 10 subjects 5 trials with the left hand followed by 5 trials with the right hand. The mirror was in the 'center' position throughout. The results for the two groups of subjects, averaged by the method used in the writer's previous cross education work, show marked intra-organic transfer. For time the averages (to the nearest whole num-

ber) are: first 5 trials, (practice) 255, 135, 96, 84, 74 seconds; second 5 trials (transfer) 96, 81, 73, 78, 68 seconds. The corresponding error averages are: (practice), 54, 30, 33, 31, 26 errors; (transfer), 38, 31, 27, 25, 25 errors. By the previously used method of computation (2, 151) the percentage of positive transfer is 88 percent for time and 60 percent for errors.

Discussion

An attempt to generalize the conditions under which negative transfer occurs has been made by the association psychologists. The statement of the principle most suited to experimental verification has been made by Bruce, to the effect that negative transfer occurs when an attempt is made to learn a new response to an old stimulus (1). At first sight this does not seem to apply to our data. The pattern of light waves is different for each mirror position, while the 'response' (tracing the maze) seems identical for all mirror positions. I believe, however, that if we can rid ourselves of the 'experience error' (4, 97) in Bruce's hypothesis, the negative transfer found in our data may show unexpected relationships with the work of earlier experimenters. The essential flaw in this S–R reformulation of the associationist position is the implicit assumption that the subject is responding to what Koffka calls the 'geographical environment' (4, 31–53), as contrasted with the 'behavioral environment.' In terms of the geographical environment, the change in mirror position is indeed primarily a stimulus change, and according to the converse aspect of Bruce's formula should lead to positive rather than negative transfer. But as an object in the behavioral environment the maze appears little different whether seen on the right, left, or in front. In fact, a mere shift in position, even when coupled with a rotation of the maze through 180 degrees, *would* cause little change in ability to trace. If the mazes were 'really' what they appear to be, the transfer might well be close to 100 percent positive.

On the other hand, however, the change in the geographical environment (by disturbing the relation between distant and proximal stimuli) necessitates the development of a new response to closely similar behavioral objects. The negative transfer results from the fact that, due to a change in the geographical environment, the response is now wrong and must be reorganized. The subject must learn a new response, not to an old stimulus in the geographical sense, but to a behavioral environment that appears to the subject unchanged in any essential that should offer difficulty in tracing. When the pattern is on the left, the subject must learn to correct a tendency to pull to the right. When the mirror position is changed and the pattern is now seen on the right, the previous learning interferes

with the task of correcting a tendency to pull to the left. And learning to correct both these tendencies interferes with the attempt to correct the front-back reversal of response in the frontal mirror position. Thus the negative transfer in our work is only a special case of the general principle. The apparent discrepancy between our results and those of other experimenters finding negative transfer, is seen to be due to the fact that in their investigations the behavioral and geographical environments maintain a constant relation, whereas in ours they do not.

One point of far-reaching importance, however, remains to be discussed: the relation of Bruce's law to learning proper. For, with constant mirror position, the appearance of the maze as a part of the behavioral environment is even less changed from trial to trial than it is from one mirror position to the next. Since one need only inspect the records of successive tracings to see that the response changes grossly from trial to trial, our formula (negative transfer from the attempt to acquire a new response to an old stimulus) would lead us to predict negative transfer rather than positive learning from one trial to the next. How then do we account for the positive transfer? The solution is that the behavioral environment (in this case in close correspondence with the geographical) is reflecting every shading and graduation of response. Every movement of the pencil causes a change in the path seen in the maze, and these changes are not incidental but essential to learning. They serve to guide the response, and without them (as Thorndike has shown) improvement is impossible (**5**).

It is thus quite consonant with (though of course not proved by) our results, that with a constant behavioral environment one could only fix more firmly (or break down) the existing response, and that learning to respond to a constant behavioral environment might in all cases give negative transfer to any other situation demanding the learning of a novel response to the same behavioral objects (**4**, 547).

It will be noted that our hypothesis does not require that transfer with shift in mirror position be always negative. Negative transfer might be minimized, absent, or go over into positive transfer under various degrees of two sorts of conditions. (1) Subjects might develop responses that would not have to be radically changed with shift in mirror position. We have already envisaged such a possibility in attempting to explain the positive transfer for subjects in the Piers-Borden experiment using the kinæsthetic method. (2) Since as behavioral environments the three mirror positions are not identical, subjects might eventually develop a differential response to each. As we have suggested in discussing the absence of negative transfer with the time measure from position 1 to position 3, it is possible

that such differentiation is beginning to make its appearance in our data. For further investigation three lines of attack suggest themselves. One might study the effect of: (*a*) more trials at each mirror position, (*b*) continuing the cycle of mirror positions (LCR, LCR, &), (*c*) generalization and training in methods.

Summary

Measured in number of errors, tracing an irregular pattern seen indirectly in a mirror placed in one of three positions (center, left, right) resulted in marked negative transfer to a second and still more marked negative transfer to a third position of the mirror. With the time measure there was negative transfer from position 1 to position 2, but equal speed on position 1 and position 3. A control experiment in cross education showed marked positive transfer, indicating that the negative transfer is a function of the shift in position of the mirror, and is not specific to the particular maze pattern.

In order to harmonize Bruce's hypothesis (that negative transfer is a function of learning a new response to an old stimulus), the hypothesis was reformulated in Koffka's terminology: negative transfer occurs in learning a new response to an old behavioral environment.

(Manuscript received February 6, 1941)

REFERENCES

1. Bruce, R. W., Conditions of transfer of training, *J. exp. Psychol.*, 1933, **16**, 343–361.
2. Cook, T. W., Studies in cross education. V. Theoretical, *Psychol. Rev.*, 1936, **43**, 149–178.
3. Hall, B. E., Transfer of training in mirror tracing, *J. exp. Psychol.*, 1939, **25**, 316–318.
4. Koffka, K., *Principles of gestalt psychology*, N. Y.: Harcourt, Brace, 1935.
5. Thorndike, E. L., *The fundamentals of learning*, N. Y. Bur. Pub., Teachers Coll., Columbia Univ., 1932, 1–32.

AN INEXPENSIVE MEMORY DRUM

BY HAROLD SCHLOSBERG

Brown University

In certain topics in the laboratory course in experimental psychology it is desirable to have all students carry out the same experiments simultaneously. This is particularly true in memory experiments, for many of the generalizations to be drawn must be based on class averages. But the problem of obtaining enough apparatus to equip each pair with a set-up is a difficult one with the average budget. In the past I have used hand-operated cardboard devices to expose memory material, for I could locate no adequate mechanical drum under $30.00. This year I discovered that a very satisfactory drum can be based on a 'stepping relay.'

The Guardian Electric Manufacturing Co., 1621–27 W. Walnut St., Chicago, Ill., will furnish a 24-step relay (S–20) especially equipped with a 3 1/8 inch diameter drum, for $5.00. See Fig. 1. A ratchet (*A*) advances the drum (*B*) through 1/24th of a revolution each time the circuit is briefly closed through the 110 volt solenoid (*C*). All that is necessary to convert the relay into a memory drum is to screw it to a wooden base, and clip on a cardboard shield to cover the unexposed words. Up to 24 syllables can be printed on a 10-inch slip of paper and fastened to the drum with a paper clip. The device can be driven by a seconds clock or a metronome.

The action of the mechanism is more powerful than necessary for rotating the light drum, and is therefore somewhat noisy. It can be quieted considerably in a half hour by the following steps:

1. Insert a felt pad to act as a stop for the armature (*E*). The back plate (*F*) on the solenoid carries a steel rod projecting into the solenoid to act as a stop. By removing two screws this plate can be taken off. A roll of felt (cut from a hat brim) should be inserted in the solenoid, and the back plate replaced with the metal rod sawed off, or simply projecting to the rear.

2. Slip a piece of rubber tubing over the bracket (*G*) that acts as a stop for the driving pawl, and readjust the position of the bracket.

3. Stretch the recoil spring (*H*) so that it will give a slower recoil. Make certain that the spring is attached to the lever (*I*) that carries the pawl. The relays are usually furnished with the spring fastened to the armature by a link, which permits the lever to buzz when operated on A.C.

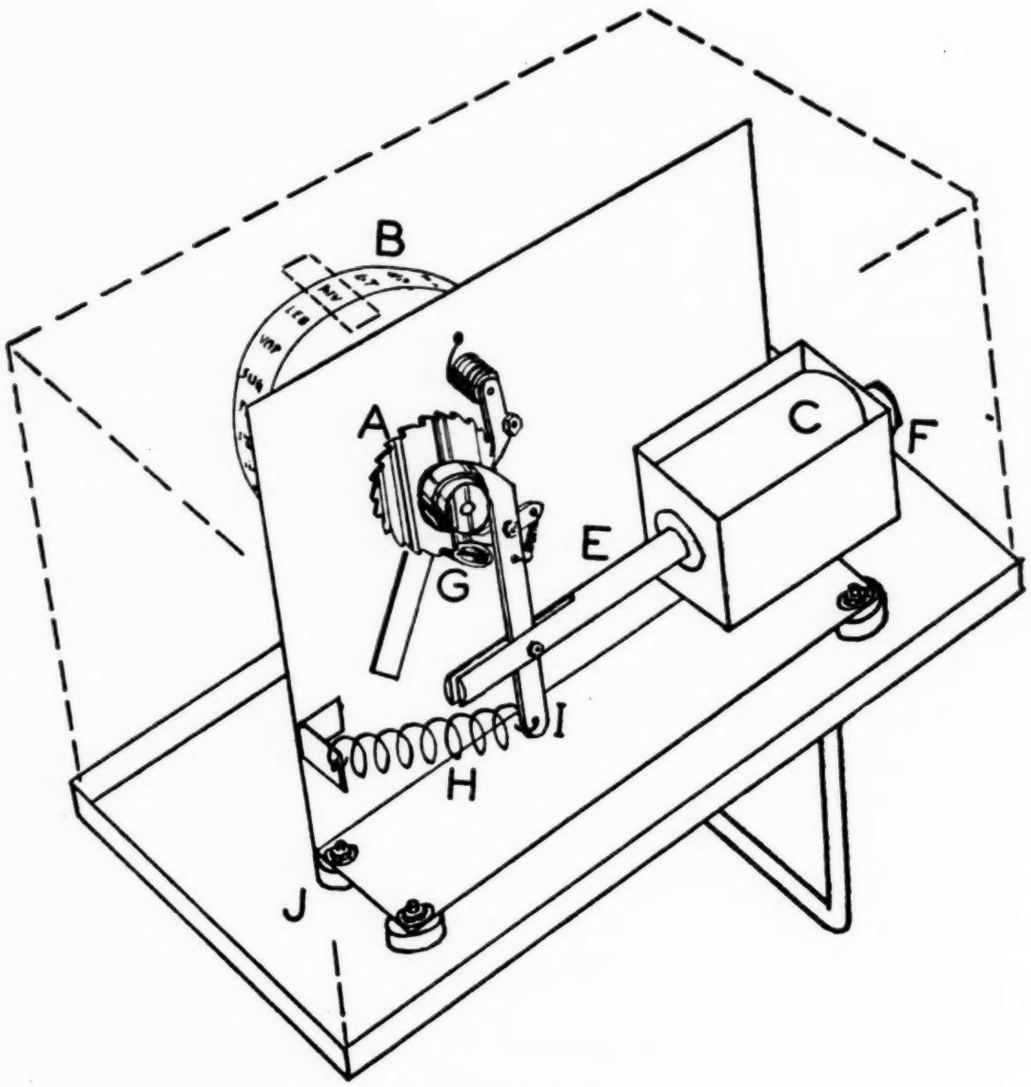

FIG. 1. For description, see text.

4. Insert a variable resistor of about 200 ohms in series with the coil.

5. In mounting the relay on its base, use rubber grommets and washers (*J*).

The unit can be given a professional appearance by mounting it in a 5 × 9 × 6 black metal box, of the type used for radio equipment. A slit 5/16 × 1 1/2 inches should be cut in the cover as an aperture. If desired, the box can be set at a convenient angle on a book rest or wire merchandise easel. The final adjustment should be made with the instrument set at the desired angle. It will operate positively with no more noise than the Ranschburg memory apparatus. The parts for such a unit cost between $7.00 and $8.00. A half dozen can be assembled and adjusted in about 15 hours.[1]

[1] I will be glad to furnish more detailed instructions, including a list of parts required and a template to locate the screw holes. I have had some slips printed to fit the drums, and will furnish samples, or larger quantities at cost ($.15 per hundred) while my surplus lasts.

For student use it is best to have mimeographed or printed slips lined to receive the syllables. The slips should have 48 spaces in 10 inches (the circumference of the drum) and some overlap. This allows ample space for 24 syllables, with a blank space around each to prevent double exposure. If the slips are uniform it is possible to use a properly-placed guide line scratched on the drum as a reference point so that the words will always be centered in the aperture without further adjustment.

At the present moment I have 7 of these drums operating on one circuit, controlled by a seconds clock (Arthur Thomas Co.) and a small 110 volt relay. Each unit is equipped with a separate line switch, so it can be stopped or started independently of the others. The drums have been completely satisfactory in the hands of inexperienced students. The only change indicated by this year's experience is a modification of the timer to increase the exposure from one to two seconds per syllable, to allow more time for the student to anticipate the syllable.

(Manuscript received February 3, 1941)

DISCUSSION

CRITICAL COMMENTS ON HIBLER'S PRESENTATION OF HIS WORK ON NEGATIVE AFTER-IMAGES OF HYPNOTICALLY INDUCED HALLUCINATED COLORS

BY ELIZABETH M. ERICKSON, B.A.

Eloise Hospital, Eloise, Michigan

In the June, 1938 issue of this JOURNAL Erickson and Erickson (5) presented experimental findings demonstrating that, by use of an adequate hypnotic technique, subjects could hallucinate colors which were invariably followed by the appropriate negative after-image. These subjects were shown to be naive concerning the theory and nature of after-images by both word-association tests and extensive questioning.

In the July, 1940 issue of this JOURNAL, Francis W. Hibler (6) reports similarly directed work which purports to contradict the work of Erickson and Erickson, and submits the following theoretical position: "since hallucinated colors definitely do not have the properties of actual colors, and since no evidence of the existence of the hallucination was discovered except a mere verbal agreement with the hypnotizer, hypnotic hallucinations are probably verbal in nature rather than sensory or a function of the central nervous system."

It is the purpose of this paper to demonstrate that Hibler's negative results derived not from the inability of a hypnotic subject to obtain true hallucinations meeting every subjective criterion of reality, but rather, from the inadequate hypnotic technique that was employed; that Hibler's results can be dismissed as complaisant coöperation on the part of inadequately hypnotized, inadequately instructed subjects, sophisticated as to the subject matter of the experiments; that the findings of Erickson and Erickson's earlier experiment as to the validity of the hallucinations and the spontaneity of the negative after-images have not been contradicted; and that Hibler's work does not constitute an experimental checking of their findings.

The first criticism to be made of Hibler's experiment concerns his selection of subjects. Of four subjects, three were sophisticated concerning after-images; the fourth had faulty knowledge, believing that in the experimental set-up for negative after-images, positive ones would appear. As will be shown later, such sophistication on the part of the subjects seriously militates against the reliability of the experimental results and precludes their adequate evaluation.

The second general criticism which may be made is that Hibler invoked his hypnotic trances too rapidly, although he notes that Erickson suggests a trance induction of at least 15 to 30 minutes. Hibler admits that the 'exact importance of the time element is certainly worthy of further research'; yet he spent only a total of eight minutes, five in verbal trance suggestion, three in which the subject simply waited, in order to induce his

hypnotic trance. In this short time he proposed to produce sufficiently profound dissociation to permit such difficult phenomena as valid visual hallucinations and possible neuro- and psycho-physiological changes; and, in such a trance, to test accurately the ability of the subject to demonstrate these phenomena. My co-worker, in his extensive experience as evidenced in work involving similar difficult phenomena, and possible physiological changes by hypnotic suggestion (**1**, **2**, **3**, **4**, **7**), has come to the conclusion that superficially a subject, especially a trained subject, can, within a period of a few seconds, manifest the appearance of a deep trance, and can easily meet all the usual criteria of the hypnotic trance. However, to establish a state adequate for serious experimental work, and so far removed from the normal waking state that hallucinations can exhibit all qualities of reality, requires that the subject be given at least 15 to 30 minutes to permit the development of possible neuro- and psycho-physiological changes and to adjust himself.

Hibler does not seem to realize that the appearance of hypnotic phenomena and the meeting of the usual trance criteria have to do only with the manifestations of a simple trance and not with highly complex phenomena to be developed out of that trance state. To establish a hypnotic trance state adequate for the induction of significant alterations of customary processes of behavior requires a period of time directly proportional to the complexity of the alterations desired. A parallel might be drawn in relation to the depth and stability of the state of physiological sleep immediately upon falling asleep and at later intervals during the course of that sleep. Hibler desired profound alterations in behavior processes, yet made no provision for the temporal requirements for such developments.

To proceed to the specific experimental set-up, it should first be noted that Hibler's preliminary control series consisted of a test to check the extent of the subject's knowledge of after-images. It consisted of questioning the subject as to his opinion of what would appear following fixation on each of the primary colors. This test may be objected to on the grounds that it possibly served to inform the subject that the experiment to follow would concern after-images, of which he did have knowledge. Erickson and Erickson's subjects had no such preliminary hints, all investigation previous to the actual test work being concealed in the form of word-association tests, and questioning being resorted to only after the experimental work had been finished.

After this preliminary series, Hibler proceeded to his tests of trance hallucinations. His apparatus set-up, consisting of a Dockeray tachistoscope using two blank grey cards so presented that the subject did not need to shift his ocular fixation, is excellent. The first card was always described to the subject as having on it a disc of one of the primary colors; then he was asked to describe what he observed on the second card.

Three of the four subjects consistently described discs of a color appropriate to the negative after-image of the primary color named. The fourth subject equally consistently described discs of the same color as the primary color named. All results were in accord with the subjects' previously expressed expectations.

Now these results may be explained in three different ways:

1. Hypnotic subjects are capable not of actual hallucinations but only of complaisant coöperation. This is the point of view to which Hibler inclines.

2. The hypnotic subjects all obtained actual hallucinations. Three subjects obtained actual negative after-images from their hallucinations. The fourth subject's hallucinations, as is so often the case, were slow to appear and develop. They had just started with the first card and attained full development with the second.

Inasmuch as the fourth subject expected after-images of a positive nature, this explanation is unlikely to be true. Nevertheless it must be considered. Had subjects been chosen who were completely naive as to the theory and nature of after-images, this difficulty could not have arisen.

3. Hypnotic subjects in a deep trance may or may not be capable of actual hallucinations. However, hypnotic subjects in a light trance, given a visual test in which they expect a definite result, may not experience visual hallucinations when so instructed, but may, at a level of verbalized agreement, give the expected answer.

Any one of the explanations is possible. The third we hold to be the most acceptable.

Hibler's next series is to be criticized not only in its interpretation, but also from the point of view of the specific instructions given to the subject.

Hibler's purpose was to test the subject for the after-images in 'the waking state, without knowledge of the hallucinated color stimulus,' to a color hallucinated in the trance state. Therefore he desired to have his subjects in a trance state while the original color was hallucinated, and awake when the after-image, if any, developed. He attempted to do this by giving the following suggestions while the subject fixated the hallucinated color disc:

"'Now I am going to wake you to the count of three. When I say *three* you will be wide awake, feeling fine in every way. While you are awakening your eyes will remain glued on the dot on the card in front of you, and you will not remember anything that happened while you were asleep. Ready, one-two-three. What do you see?' (Experimenter removed card with hallucinated color between count of one and two.)"

Now, using the same experimental set-up and suggestions, but using real rather than hallucinated color discs, each subject had reported the appropriate negative after-image in every case, showing that the awakening from trance alone does not disrupt the actual development of a real after-image. In using real rather than hallucinated cards, the experimenter was depending upon a reality situation totally independent of the hypnotic state. Memory of having seen the real card does not enter into the development of a negative after-image. Thus, he could be shown the real card in the trance state, awakened with no memory of having seen it and yet, in the ordinary course of events, experience the negative after-image.

But when hallucinated colors were used in this series, no after-image appeared in any case. Analysis of the suggestion given shows that the subjects had been told to hold themselves in readiness to awaken. Now

this suggestion in itself served to arouse the subject from the hypnotic trance to a certain extent, putting him into a ready-to-awaken or an altered trance state, more akin to the waking than to the deep trance state. Essentially, the subject has been asked to hallucinate in the trance state, is then asked to awaken, and thus to abolish the trance state, and yet, at the same time, to continue his trance activities. The experimental data obtained from subjects in this borderline state relates to the realities to be expected after awakening, and not to the activities of the trance which have been abruptly and unintentionally abolished. Thus, Hibler requests certain responses and then inhibits them by changing the total situation upon which those responses depend. Yet, Hibler gives as "the only possible interpretation that if an after-image of any sort is experienced in the trance state as the result of an hallucinated color, it must be very different from that experienced from the presentation of an actual color stimulus."

Hibler's final series was much more complicated. It assumes the following: (p. 55) "If the subject actually experiences red, it should make no difference if that red is super-imposed over blue, or is on a grey card."

This primary assumption is certainly disputable. An hallucination in the usual sense is without doubt superimposed upon a background of reality. The subject does not see the hallucinated red disc all by itself with the rest of the visual field a neutral haze; he sees it in the appropriate opening in the tachistoscope. If that opening in the tachistoscope is occupied by a blank grey card, his task, to hallucinate a red disc on it, is relatively easy. But if it is occupied by a grey card on which there is already a blue disc, the difficulty of his task is increased many fold; it now becomes that of seeing a distinct object, the blue disc, as one totally different, a red one; and this transformation is to have absolute validity.

Now how does Hibler propose to suggest to his subject that this extremely difficult task be accomplished? He gives the following suggestion:

"You see this disc, what color is it? (Pointing to the red disc.) That's right. Now this disc (Blue) is exactly the same as this one (Red). It is the same in every way, the same hue, brightness, and saturation. It is the *same* color. Now take this card in your hand and describe this disc to me. (Subject takes blue disc and describes it as red, giving size and shade as a rule.) That's right, now remember, this (Blue) is the same as this (Red) until I tell you differently."

Before commenting on these specific suggestions, it is necessary to emphasize the care which must be taken with the wording of hypnotic suggestions. It is of extreme importance in giving hypnotic suggestions, particularly in what is intended to be a closely controlled experimental situation, to be certain that the meaning of every word the experimenter says has been carefully evaluated, especially for its most literal sense, so that the subject is not led astray by the superficial meanings of the words employed, but, rather, fully comprehends the actual significance of the instructions.

In experimental work with waking subjects, this extreme care is not ordinarily so necessary. For example, a subject was given an examination which necessitated filling out a questionnaire concerning her age, nativity, and education. One question ran 'Where born: City————————

State————————.' In the waking state the subject unhesitatingly filled in the name of the city and state. The same subject in the hypnotic state wrote 'Yes' in each blank. This is not a chance example, but rather one typical of numerous comparable instances.

Similarly, a subject was shown a tray containing a number of objects for a few moments and asked what he saw. In the waking state it made no difference which of these two forms of the question was used:

"Will you please tell me everything you saw on the tray?"

"Please tell me everything you saw on the tray."

Yet, when hypnotized, the same subject usually answered simply 'yes' to the first form of the question and quietly awaited further instructions, but would give the expected type of answer to the second form.

In addition to this literalness, the hypnotized subject often responds to cues which the waking subject would not notice or would ignore as accidental or insignificant. One needs only to see a subject in a deep trance state to realize and be impressed by the fact that the subject can, upon observing a slight gesture, an intonation, a change in wording on the part of the hypnotist, which may even have been unwittingly made, unhesitatingly carry out the suggestion implied by it.

To return now to Hibler's suggestions, the final sentence, "That's right, now remember, this (Blue) is the same as this (Red) until I tell you differently," distinctly told the subjects that the change of color was to be one of understood agreement only and thus effectively destroyed any force which the preceding suggestions might have had. It is not sufficient to have the subject agree only—rather the subject must perform the infinitely harder task of experiencing, and the acceptance of Hibler's suggestions did not imply an acting upon them but only an abidance by them. To show the inadequacy and self-contradictory character of any such suggestions, one might parallel them as follows: "You see this watch. It is a gift to you. It is to be all yours, your very own, permanently yours, *until I tell you differently.*"

But even without this final nullification, the suggestions in themselves are far too simple, brief, and casual to cause such a marked effect as was desired. Adequate suggestions should have been similar to the following, with sufficient time allowed for full comprehension of and full response to each suggestion and a variation from one subject to the next to insure individual understanding on the same significances:

"You see this disc. What color is it? (Red). I want you to see it plainly and clearly and to continue to see this red disc plainly and clearly. Now I want you to see this other disc. What color is it? (Blue). Now I want you to look at it carefully and steadily because slowly, gradually, sooner or later it will lose its present color and will acquire a new, different and equally bright color. And as you keep looking you will begin to note its changing; you will become increasingly aware of the change and at first, as you note the change, this disc becomes increasingly neutral in appearance. Then, after it becomes neutral it will acquire a new and bright and familiar color, and as you begin to wonder what this color may be, you will begin to note that this disc is rapidly becoming red, even as red as the first disc; and

as you look at it you see and feel and know the redness, just as you see and feel and know the redness of the first disc, and so now you will continue to see either of these red discs at any time necessary, seeing them both clearly, fully, and readily."

The purpose of the suggestions is to induce an experiential response, not a tacit agreement. This is difficult; it requires a subject in a deep trance; it requires repeated suggestions making the desired change in color through small gradations. The color is first faded to neutral because that is a change more acceptable to the subject, in keeping with his general past experience, such as scotopic vision. The experiential values of the first red disc are emphasized, and finally of the two red discs, and after the change is made, no word is spoken which suggests or even hints of pretense or of any difference between the two. Then and only then can we expect genuine responses.

Hibler's results are exactly as would be expected: when results were reported in the waking state (the 'hypnotic' work having been done in a state preparatory to awakening), all after-images were of the real rather than the suggested color; when results were reported in the trance state, two subjects reported the results they expected to get from the suggested color (positive on the part of one, negative on the part of the other), while two reported the after-images of the real rather than the suggested color. In other words, all these results can be readily explained on the assumption of mere verbal coöperation on the part of the subjects, and failure to induce experiential responses.

In Erickson and Erickson's study, the subjects, so far as could be determined by association tests and extensive questioning afterwards, and from their selection from the classes of commerce and engineering freshmen, were naive concerning the nature of complementary colors. Well-trained subjects were used, a period of 30 to 45 minutes of trance induction given to insure profound trances, and experiential responses rather than verbalized agreement. The use of alternate sheets of white paper, described as all being colored, the subject and the experimenter to name the color of the alternate sheets, prevented the subject from even realizing that the test concerned after-images, or (since they were found to know nothing of after-images), of a white sheet of paper seen after looking at a colored one. Of five subjects, one failed completely to meet the test situation apparently for the reason that he could not hallucinate colors. The other four invariably described the intervening white sheets as of the complementary color appropriate to the hallucinated color of the preceding sheet. Hibler's carefully set-up apparatus gives the superficial impression of greater accuracy than the simple sheets of white paper; yet careful evaluation showed that his extreme control of experimental conditions concerned everything but the actual essentials.

Hibler's experiment proved only that an inadequately hypnotized subject who knows about after-images, sees the appropriate after-images in what he realizes to be an after-image set-up; that subjects instructed to hallucinate colors in a semi-waking state do not experience after-images of those colors; and that subjects instructed to regard and treat one color as another, know-

ing about after-images, may or may not carry the coöperation so far as to report the after-image they assume to be correct.

Hence, Erickson and Erickson's earlier experiment, not having been repeated or even checked in any detail by Hibler's, must still be considered to be accurate, and its final conclusion to stand: that the hallucination of various colors by hypnotic subjects can be sufficiently valid subjectively to permit a spontaneous and invariable hallucination of complementary colors somewhat in the form of negative after-images.

The real contribution of Hibler's experiment to hypnotic investigative work is its demonstration of how misleading and unreliable experimental results can be when their acceptance is based upon the completion of an experimental procedure without due regard or provision for the nature and character of the phenomena under investigation by that procedure.

(Manuscript received January 22, 1941)

REFERENCES

1. Erickson, M. H., Development of apparent unconsciousness during hypnotic reliving of a traumatic experience, *Arch. of Neur. and Psychiat.*, 1937, **38**, 1282–1288.
2. Erickson, M. H., A study of clinical and experimental findings on hypnotic deafness: I. Clinical experimentation and findings, *J. gen. Psychol.*, 1938, **19**, 127–150.
3. Erickson, M. H., A study of clinical and experimental findings on hypnotic deafness: II. Experimental findings with a conditioned response technique, *J. gen. Psychol.*, 1938, **19**, 151–167.
4. Erickson, M. H., The induction of color blindness by a technique of hypnotic suggestion, *J. gen. Psychol.*, 1939, **20**, 61–89.
5. Erickson, M. H., and Erickson, E. M., The hypnotic induction of hallucinatory color vision followed by pseudo negative after-images, *J. exp. Psychol.*, 1938, **22**, 581–588.
6. Hibler, F. W., An experimental investigation of negative after-images of hallucinated colors in hypnosis, *J. exp. Psychol.*, 1940, **27**, 45–57.
7. Huston, P., Shakow, D., and Erickson, M. H., A study of hypnotically induced complexes by means of the Luria technique, *J. gen. Psychol.*, 1934, **11**, 65–97.

NOTE ON MRS. ERICKSON'S COMMENTS ON HIBLER'S WORK ON NEGATIVE AFTER-IMAGES OF HYPNOTICALLY INDUCED HALLUCINATED COLORS

BY FRANCIS W. HIBLER

Mrs. Erickson (**1**) has written a careful and challenging critique of my work on negative after-images of hypnotically induced colors (**7**) and has also pointed out contrasts in technique, as well as results, with Erickson and Erickson's earlier work in the same field (**6**). I appreciate her comments as they may stimulate further fruitful research. Each point raised invites extensive discussion, but as the space is limited, I shall attempt to make my remarks as brief as possible.

Of all the various points on which Mrs. Erickson directs her attack, the one in which she charges that my subjects were inadequately hypnotized appears to me basic. If this charge be true, it must be granted that all the

data obtained are invalid, and further discussion of the paper would be fruitless. This attack is based on the following evidence (1): Dr. Erickson:

> "in his extensive experience as evidenced in work involving similar difficult phenomena, and possible profound neuro- and psycho-physiological changes by hypnotic suggestion (1, 2, 3, 4, 7) has come to the conclusion that superficially a subject, expecially a trained subject, can, within a period of a few seconds, manifest the appearance of a deep trance, and can easily meet all the usual criteria of the hypnotic trance. However, to establish a state adequate for serious experimental work, and so far removed from the normal waking state that hallucinations can exhibit all qualities of reality, requires that the subject be given at least 15 to 30 minutes to permit the development of possible neuro- and psycho-physiological changes and to adjust himself."

A review of the above-cited articles, however, fails to reveal any carefully controlled experiment that yields objective data on which Dr. Erickson bases his conclusion. All investigators are especially interested in obtaining objective criteria of a deep trance, and there is a genuine need for the acceptance of a 'standard technique of inducing hypnosis and measuring its depth' as suggested by Young (12) in his recent review of experimental hypnotism. (P. 95.) If this induction always requires 15 to 30 minutes, as asserted, then it should be demonstrated by a carefully controlled experiment that makes a comparison between the results obtained in an extended period of induction versus those obtained in a short period. Inasmuch as Dr. Erickson's belief in this matter is contrary to the accepted practice of many investigators in this field, some objective evidence gathered under rigid scientific control would be of great value.

Hull (8), for example, in a chapter entitled 'Elementary Phenomena' describes phenomena characteristic of deep hypnosis as: hallucinations, negative hallucinations, hypnotic analgesia, post-hypnotic amnesia and post-hypnotic suggestion. He concludes that a subject showing the above phenomena "represents about as profound a trance as is ever obtained." (P. 30–31.) Each of my subjects satisfied Hull's criteria in entirety. Hence, until objective evidence supporting Dr. Erickson's contention is produced, great care should be exercised in drawing dogmatic conclusions about this matter. Space forbids a discussion here of some of the interesting theoretical implications of Dr. Erickson's hypothesis.

Mrs. Erickson also raised the question of the relative merits of our two investigations as to scientific control and experimental procedure (1). Inasmuch as the technique in my investigation has been accurately presented, together with critical comments by an experienced investigator in the field, it may be fairly evaluated. It will be noted that my paper (7) *explicitly stated* the conditions of the experiment and it must stand or fall on its merits. Certain omissions in the original Erickson paper (6), however, makes it most difficult to compare the adequacy of the techniques of the two investigations.

It will be remembered that the Erickson technique employed the use of alternate sheets of white paper, described as being all colored, the experimenter naming the color of the first sheet and the hypnotized subject the color of the next, and so on through a series of 16 sheets.

Some of the questions that arise are as follows:

1. Was any attempt made to validate the use of word association as a technique of determining the naivete of their subjects? Rosanoff (11), for example, finds that out of 1000 subjects, selected at random, 99 responded to the stimulus word of 'red' by saying 'blue,' while 30 responded with 'green,' and 254 said 'color.' Hence, one wonders what results were obtained on the association series, that an unequivocal conclusion that Erickson's subjects were naive.
2. There is little description of any controls used in the presentation of the sheets of paper on which the colors were hallucinated. One wonders, for example, as to the tempora duration of the presentation of the papers (even if it was not constant), and expecially as to *what time interval*, if any, was employed between the naming of the colors on the alternate papers. This is not stated in the report.
3. There is no mention of any attempt to validate their technique to determine whether or not they would get the same results with *actual* colors in the *waking* state, that they reported with *hallucinated* colors in the *trance* state. It would appear that this should be the absolute minimum in the way of scientific control.

In studying Mrs. Erickson's comments one must be impressed again with the fact that scientific progress can be made only under the most rigid scientific experimentation. For her, any investigator who does not accept the Erickson assumptions as to trance induction and suggestion, thereby fails to allow for 'possible neuro-psycho-physiological changes' in hypnosis and hence produces results that are 'misleading and unreliable' as they are without 'due respect or provision for the nature and character of the phenomena under investigation.' Does not Mrs. Erickson, in making this statement, assume that she already knows the 'nature and character' of the phenomena that we are purportedly investigating?

As Pattie (10) has pointed out recently after reviewing the literature on reported cases of hypnotically induced blisters: "The observers have looked for blisters. . . ." He then goes on to suggest that more rigid scientific controls must be employed before a conclusion is justified. Likewise, in the area of hypnotically induced hallucinated colors, each investigator must be extremely careful that some *a priori* assumption as to the nature of hypnosis does not influence his concept of what constitutes an adequate experimental technique.

(Manuscript received February 3, 1941)

REFERENCES

1. Erickson, E. M., Critical comments on Hibler's presentation of his work on negative after-images of hypnotically induced hallucinated colors, *J. exp. Psychol.*, 1941, **29**, 164–170.
2. Erickson, M. H., Development of apparent unconsciousness during hypnotic reliving of a traumatic experience, *Arch. Neur. and Psychiat.*, 1937, **38**, 1282–1288.
3. Erickson, M. H., A study of clinical and experimental findings on hypnotic deafness: I. Clinical experimentation and findings, *J. gen. Psychol.*, 1938, **19**, 127–150.
4. Erickson, M. H., A study of clinical and experimental findings on hypnotic deafness: II. Experimental findings with a conditioned response technique, *J. gen. Psychol.*, 1938, **19**, 151–167.
5. Erickson, M. H., The induction of color blindness by a technique of hypnotic suggestion, *J. gen. Psychol.*, 1939, **20**, 61–89.
6. Erickson, M. H., and Erickson, E. M., The hypnotic induction of hallucinatory color vision followed by pseudo-negative after-images, *J. exp. Psychol.*, 1938, **22**, 581–588.
7. Hibler, F. W., An experimental investigation of negative after-images of hallucinated colors in hypnosis, *J. exp. Psychol.*, 1940, **27**, 45–57.
8. Hull, C. L., *Hypnosis and Suggestibility*, D. Appleton-Century Company, New York, 1933.

9. Huston, F., Shakow, D., and Erickson, M. H., A study of hypnotically induced complexes by means of the Luria technique, *J. gen. Psychol.*, 1934, **11**, 65–97.
10. Pattie, F. A., The production of blisters by hypnotic suggestion, *J. abnorm. (soc.) Psychol.*, 1941, **36**, 62–72.
11. Rosanoff, A. J., *Manual of Psychiatry*, John Wiley & Sons, London, 1920.
12. Young, P. C., Experimental hypnotism: a review, *Psychol. Bull.*, 1941, **38**, 92–104.

A CRITICISM OF LEWIS AND LARSEN'S 'MASKED AUDITORY THRESHOLDS'

BY RALPH H. GUNDLACH AND ROBERT DICE

University of Washington

In the recent report of an excellent study by Don Lewis and M. J. Larsen entitled Measurement of masked auditory thresholds (This Journal, 1940, **27**, 601–623), the authors either make a considerable error, or else almost completely mislead the reader in their graphical presentation of data in that part of Figs. 8, 9, 10, 11 and 12 with abscissæ labeled 'sensation level' or 'masked sensation level.'

A masked threshold, Lewis and Larsen point out, is the threshold of a tone determined in the presence of a background of one or more masking tones. In the experiments which they describe, the masking tones and the tone whose threshold was in question were all members of the same harmonic series, since the authors were interested in the question of timbre.

In the presentation of their material in Figs. 1, 4, 5 and 6 (and similar parts of the remaining figures), the authors employ ordinates and abscissæ that, in dealing with sounds, treat them in strictly physical terms; *i.e.*, in terms of frequency and intensity level. The important item is intensity level, which is a measure of the physical energy in db above 10^{-16} watt as applied by the electrical equipment to the subject's ear. This is of course not only perfectly proper, but necessary for identifying the conditions within which the experiment proceeded.

Lewis and Larsen point out, however, that at least in the case of threshold determinations for 'harmonics VI and VII' they met with the confusing presence of what they found to be a subjective tone of the same frequency as the harmonic in question, and with a loudness equal to a stimulus-imposed intensity level of 30 or 32 db (p. 613).

It is the presence of this subjective harmonic, and the vacillating attitude of sometimes-considering, sometimes-ignoring it by Lewis and Larsen, which constitutes the basis of our criticism. In their determination of thresholds, intensity level, sensation level, and masked sensation level, the authors in every case employ figures which completely ignore the presence of this rather strong subjective harmonic they were at such pains to point out and to measure. It is of course quite important to indicate in physical terms the readings of the apparatus, both as regards intensity level and the threshold variations, as they properly do in all their figures with an 'intensity level' abscissa.

But when they talk of 'sensation level' and 'masked sensation level' then a new base of reference is introduced. Now the question has to do with the phenomenal tone and its loudness, whether considerable or just perceptible. As Lewis and Larsen point out, the sensation level of any sound is the amount (in db) by which its loudness exceeds its intensity at the threshold of audibility. They have deliberately defined the Masked Absolute Threshold in terms of the imposed external intensity level, ignoring the subjective intensity of the masked harmonic; and hence, their m.a.t. is actually above the threshold of audibility. These authors have, in fact, experimentally demonstrated that they can reduce the loudness of this subjective harmonic by the presentation of a weak objective tone out of phase with it.

Our objection is this: the abscissæ which read 'sensation level' and 'masked sensation level' imply a reference to the phenomenal loudness of the tones but actually omit calculation of the subjective harmonic, which is an admitted component.[1] In this case, every point plotted against these abscissæ in all these figures is in the wrong place, and needs to be shifted, right or left. Furthermore, in the same figures, if the term 'Threshold in DB' is thought likewise to refer to the phenomenal loudness of tones, which it certainly seems to, these values also are all in error, and need to be recomputed in terms of a base which is equal to the sum of the stimulus-imposed intensity level plus the subjective loudness.

The definition of Masked Absolute Threshold, when that value is actually not the same as the just audible tone, cannot be used interchangeably with the usual meaning of the term, absolute threshold, without creating considerable confusion.

(Manuscript received April 23, 1941)

REPLY TO GUNDLACH AND DICE

BY DON LEWIS

State University of Iowa

When Professor Gundlach and Mr. Dice begin by stating that Larsen and I "either make a considerable error, or else almost completely mislead the reader in . . . part of Figs. 8, 9, 10, 11, and 12" and later add that "every point plotted against these [sensation level and masked sensation level] abscissæ in all these figures is in the wrong place, and needs to be shifted, right or left," they lead me to believe that they have not grasped the full meaning of the terms 'sensation level' and 'masked sensation level.' This belief is not weakened when they credit us gratuitously with the statement that "the sensation level of any sound is the amount (in db) by which its loudness [*sic*] exceeds its intensity at the threshold of audibility."

[1] The same criticism holds of the interpretation of Fig. 5 as expressed in the sentence starting line 11, page 610.

Their substitution of 'loudness' for 'intensity' in the definition as we worded it is, of course, not warranted.

Neither sensation level nor masked sensation level is a measure of loudness—'phenomenal' or otherwise. Both levels are expressed in terms of number of decibels above clearly defined reference intensities; and a decibel, it should be emphasized, is a unit, not in a loudness scale, but in a logarithmic scale of sound intensity. A logarithmic scale of a physical dimension does not become transformed into a psychological scale merely by virtue of its having a reference point which is based upon the responses of one or more observers. To say that sensation level or masked sensation level "has to do with the phenomenal tone and its loudness" is to use the terms incorrectly. (It might be said, in fairness to Gundlach and Dice, that perhaps a majority of psychologists fail to realize that sensation level has reference to a physical dimension and not to a psychological magnitude.)

Our critics charge us with taking a "vacillating attitude of sometimes-considering, sometimes-ignoring" the presence of subjective harmonics; and this, they say, constitutes the basis of their criticism. If they would carefully re-read our paper, and also the second and third refrences cited therein, they would discover that the measurements which we have been making (especially measurements of masked absolute thresholds) furnish data by means of which the magnitude of subjective harmonics may be estimated. One of our principal aims has been to devise and use methods of experimentation which would enable us to quantify subjective elements.

In our paper, on page 613, footnote 6, we give computed values for the magnitude of subjective harmonic VII, as generated in each of two ears under specified experimental conditions. Gundlach and Dice accept and use these values. Yet they seem to have overlooked the fact that the values could not have been estimated if we had lacked the threshold data which they seem to censure us for securing.

Many of our plotted values, they say, are in error, "and need to be re-computed in terms of a base which is equal to the sum of the stimulus-imposed intensity level plus the subjective loudness." What they apparently expected us to do was to measure masked absolute threshold values, then use these threshold values in estimating the magnitudes of subjective harmonics, and finally 're-compute' the original threshold values in terms of the results of computations made upon them. We may be guilty of some scientific sins, but not, we hope, of complete circularity.

It is true that we defined the term masked absolute threshold as the intensity level of the imposed acoustic wave which corresponded to a required psychometric condition; and it is also true that we did this deliberately. The reason was that sensory thresholds can be defined only in relation to some aspect of a stimulus.

Why confusion should arise from the use of 'masked absolute threshold' in one sense and 'absolute threshold' in another is not made clear. The Acoustical Society of America has adopted the following definition of masking: "The masking of a sound is the shift of the threshold of audibility [the absolute threshold] of the masked sound due to the presence of the

masking sound. The unit is the decibel." [1] It seems entirely legitimate to apply the term masked threshold—masked absolute threshold or masked differential threshold, as the case may be—to a threshold which is shifted because of the presence of one or more masking harmonics.

(Manuscript received May 31, 1941)

[1] *J. Acous. Soc. Amer.*, 1937, **9**, p. 64.